TRANSWORLD STUDENT LIBRARY

- Abilities
- Strategies
- The inductive method
- Methods of proof
- Problems and extensions
- Exercises
- Answers and hints

Much of the development of mathematics has evolved through the stimulation provided by searches for the solutions of problems, and mathematical knowledge itself is tested by problems, the solving of which necessitates the exercise of various abilities and skills. At a relatively low level what is required is an understanding of the problem and the correct application of some standard technique. At a higher level—say, A-level or University entrance standard—more is demanded: the ability to formulate a suitable mathematical model, and to use imagination to think of algorithms or analogies when no method of solution is immediately evident.

This book, almost unique in the field, analyses the approach to problems in mathematics and outlines a strategy for tackling them in general. While problems of the 'to find' type often depend on a standard technique for their solution, other problems normally need some inductive investigation. The inductive process is therefore discussed and the concepts expressed are illustrated with a collection of examples. Even when a solution has been found to a problem, there remains the question of proof. A whole variety of proofs—inductive, illustrative, exhaustive, etc.—are analysed, again with examples. Finally, a collection of exercises is provided, some requiring for their solution purely manipulative ability, others requiring special application or insight. Answers and helpful hints are provided.

Both students of mathematics and their teachers will derive pleasure, stimulation, and knowledge from this unusual book.

UK .85p Australia $2.45 New Zealand $2.60

* RECOMMENDED PRICE ONLY

0522 40015 7

S Moses

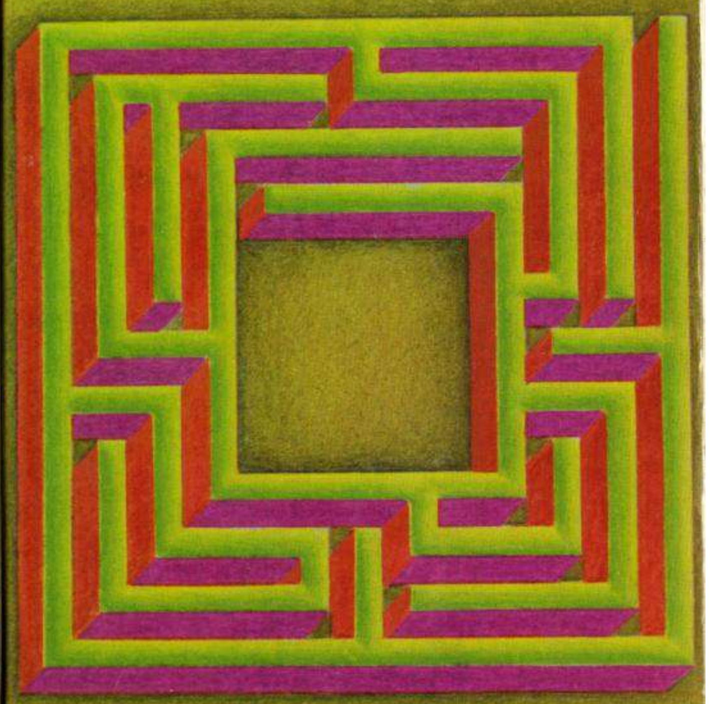
THE ART OF PROBLEM

0 522 40015 7

TRANSWORLD STUDENT LIBRARY

THE ART OF PROBLEM-SOLVING

S Moses



Transworld Student Library

General Editor
H. GRAHAM FLEGG, MA, DCAe, CEng., FIMA, MIEE, MRAeS, FRMetS

Reader in Mathematics, The Open University

Other books in this series

Boolean Algebra H. G. Flegg
Theoretical Statistics S. N. Collings
Points and Arrows: the Theory of Graphs A. Kaufmann
Meteorology H.-J. Tanck
Field Projects in Sociology J. P. Wiseman and M. S. Aron
The Unknown Ego T. Brocher
Calculus via Numerical Analysis A. Graham and G. Read
Basic Mathematical Structures I N. Gowar and H. G. Flegg
Basic Mathematical Structures II N. Gowar and H. G. Flegg
Towards Quantum Mechanics J. Cunningham
Reliability: a Mathematical Approach A. Kaufmann
Paradoxes of Physics P. Chambadal
New Perspectives in the Theory of Evolution H. Querner, H. Hölder, A. Egelhaaf, J. Jacobs, and G. Heberer
Numerical Analysis A. Graham
Biology: the Ultimate Science B. Hocking
The Art of Problem-Solving S. Moses
Aspects of Modern Geometry C. C. H. Barker
Introducing Real Analysis D. H. Fowler
The Evolution of Mathematical Concepts R. L. Wilder
Mathematics and Mathematicians I P. Dedron and J. Itard

The Art of Problem-Solving

Stanley Moses

TRANSWORLD PUBLISHERS LTD
In association with Richard Sadler Ltd

THE ART OF PROBLEM-SOLVING

A TRANSWORLD STUDENT LIBRARY BOOK 0 552 40014 9

First publication in Great Britain

PRINTING HISTORY

Transworld Student Library Edition published 1974
in association with Richard Sadler Ltd

Transworld Student Library Books are published by
Transworld Publishers Ltd, Cavendish House,
57–59 Uxbridge Road, Ealing, London, W.5

Set in Times 10/12 pt

Made and printed in Great Britain by
Richard Clay (The Chaucer Press) Ltd
Bungay, Suffolk

Contents

Foreword

by Professor R. Brown

I would like to recommend warmly this book as one which introduces the reader to two virtues very important for a mathematician: namely, curiosity and delight. The point of the book is to deal with problems and problem-solving, and to introduce the reader to methods of setting about the business of solving problems. Curiosity is the wish to see how the problem can be done, even when the standard methods fail. And delight comes in because sometimes an entirely novel way of looking at a problem or of modelling it is needed in order to solve it.

It may not be generally realized that problems are one of the chief motivating aspects of mathematics, and that it is in attempts to solve problems or to understand solutions already produced that the subject is advanced. For this reason an aspiring mathematician, or anyone who wants to understand what the subject is about, should be introduced to many problems, some easy and some difficult, and should understand and learn some of the basic tricks and techniques of problem-solving. For example, one of the standard methods is to set out the data and the conclusion clearly and to see how much of a gap is between them. You may of course only get as far as a sheet of paper with the data at the top, the conclusions at the bottom, and a large blank space between. But even this is better than a blank sheet of

paper, and it is also at this stage that the fun begins, as you start trying to find some way of filling the gap.

Solving problems is a skill, which can to a certain extent be learned, not just by reading books about methods, but by examining models of the problem-solving method, by applying them in practice, by evaluating them, re-applying and re-evaluating. So you learn that when the obvious methods fail, then it is necessary to try and twist the question into a new shape which may prove more amenable. There is an important question, but one not susceptible of a complete answer, as to how long an apparently unsuccessful method should be persisted in. It would be interesting to see how long it takes even an experienced mathematician to give up trying to prove in example 13 of Chapter 3 that $U_n - 1/n \to 0$ as $n \to \infty$ and to think instead of proving $1/(nU_n) \to 1$ or that $u_n = 1/(n+\sigma_n)$ where $\sigma_n \to 0$.

The exercise of any skill such as problem-solving requires a number of ordinary virtues such as hope and courage. Hope that some sort of success will eventually come is clearly necessary, and courage is often needed to carry through a calculation in spite of its apparent difficulty or complexity. Also many of us get carried through a difficult patch, or pick up pens to try again, out of annoyance with both oneself for not succeeding and with the problem for not coming out as one would like.

An interesting question is: How much of the methods shown in this book is applicable to those using mathematics and to research in mathematics? Indeed, there are probably many readers who will wonder how it is possible to go about doing research in mathematics at all. The answer to this question is as suggested earlier, that in order to do research in mathematics, and also in using mathematics, you start with a problem, look at methods which have been used for attacking this kind of problem, and then try and formulate some sort of strategy for attacking the

problem. So for this, and indeed for understanding and using mathematics at any level, it is useful to have assimilated and practised all the usual strategies of problem-solving.

The position of the research worker has some advantages and some disadvantages as compared with that of the person who is given a problem which has already been solved. The main disadvantage is that the research worker has no definite knowledge of the limits of difficulty of the problem: so there is a serious question of deciding how much time and effort a particular problem or a particular method is worth. As extreme examples, there are many classical and challenging problems around (for example, Goldbach's conjecture that every even integer is the sum of two odd primes, the four colour problem, Fermat's last theorem, and many others) but it would require a rare order of courage and hope, and perhaps foolhardiness, to take on one of these. But even when considering a more mundane problem, a considerable amount of judgement and experience is needed to estimate that it is a problem in which one can reasonably hope for some sort of success.

The advantages a research worker has are rather different. First of all he can choose his ground. If he does not like, or is unsuccessful with, a particular problem, he can choose a different problem, or a simpler one, or he can consider the relationship between his problem and other, possibly equally difficult, problems. What is asked of a research worker is not that he should solve a particular problem, but that over-all he should show some progress.

This brings us to the advantage the research worker has of scale. It is very helpful for a researcher to be considering a rather broad kind of problem from which he can extract lots of smaller problems which he can solve and so usefully develop his skills and confidence. It is also this interest in pushing forward an important and general area of research which often supplies the motivation for attacking a particular

problem. In this respect, there is a lot to be said for taking on a rather challenging problem, as long as the worker has some general line of approach which offers hope of some success and some advantage however small.

Thus I hope that many who are learning mathematics, for its own sake, for its applications, or for interest as a career in teaching or otherwise, will study this book as an adjunct to their courses, and will try and acquire some experience and expertise in problem-solving, and so realize that mathematics is not just an abstract contemplation of beautiful patterns, but has also its interest as a fascinating activity and craft.

Bangor 1973

R. BROWN
Professor of Pure Mathematics
University College of North Wales, Bangor

List of Symbols and Notations

$=$	equals
$\neq$	not equal to
$\equiv$	congruent
$\rightarrow$	either a mapping or approaches a limit (according to the context)
$\Rightarrow$	implies
$\Leftrightarrow$	logically equivalent
$\in$	is a member of: '$x \in S$' says 'x is a member of set S'
$\notin$	is not a member of
$\subset$	is contained in: '$A \subset B$' says 'every member of A is a member of B'
$\cup$	union
$<$	less than
$\leqslant$	less than or equal to
$>$	greater than
$\geqslant$	greater than or equal to
$!$	the factorial sign
$\parallel$	parallel
$\bar{\alpha}$	the complex conjugate of α
$\varnothing$	the empty set
$\mathbf{Z}^+$	the set of positive integers
$\mathbf{R}$	the set of real numbers

$[a_{rs}]$	the element in the rth row and sth column of a matrix
$\binom{n}{r}$	the same as ${}^nC_r = n!/r!\,(n-r)!$
$\left[\frac{n}{r}\right]$	the largest integer less than or equal to the value $\frac{n}{r}$
$[a, b]$	the closed interval $a \leqslant x \leqslant b$
(a, b)	the open interval $a < x < b$
$\sum_1^n f(r)$	the summation from $r = 1$ to $r = n$, inclusive
$\int x \to f(x)$	the Leibniz integral $\int f(x)dx$
$\sim$**a**	the proposition '**not a**' (where **a** is a proposition)

Introduction

The greater part of this book is based on two lectures, one on 'Problem-Solving' given to the North Wales branch of the Mathematical Association and the other given to the Mathematics Foundation Course Team at the Open University on 'The Inductive Process'. The chapter on 'Methods of Proof' was added because these are seldom discussed in books below university standard and the usual context is then a textbook on 'Logic'. Although an understanding of logic is essential to a complete grasp of the various methods of proof and disproof, it is surely necessary that students coming to university should have some reasonable idea of what constitutes a proof. Except for chapter 3 and those problems which are traditional the ideas are my own and have been developed over 40 years experience in teaching and lecturing at various levels. The ideas expressed in chapter 3 are undoubtedly influenced by my study in 1954 of Professor Polya's books on *Mathematics and Plausible Reasoning*. The principles discussed there have over the intervening years become so embedded in my teaching practice that I tend to regard them as traditional. After my transfer in 1960 to university lecturing I had no time to keep in touch with books on problem-solving and so I missed the publication in 1962 of the sequel on *Mathematical Discovery* at the time of its publication.

Although a mathematical student needs a capacity for logical thought, it is essential that he also acquires a reason-

able facility in manipulating symbols. It may be true that skill in such manipulation can embellish trivial ideas and that slick tricks which do not involve some general principle are not important in themselves: yet without sufficient practice in standard elementary operations the student will often be unduly handicapped in his attempts to solve even straightforward problems. In the short term the greatest benefit from a study of a given problem is the identification of certain features in the solution that seem useful in tackling similar problems, because this enables us to construct a model which will apply to these similar problems. In the long term an improvement in the ability to solve problems requiring some degree of originality will come mainly from a study of methodical work on problem-solving, and this leads to a study of the 'inductive method' explained in chapter 3. For a real understanding of the inductive process the reader should ponder over the arguments as he comes across them.

I am very grateful to Professor R. Brown for writing the preface and for many helpful comments on the manuscript of this book. I am also indebted to Graham Flegg for his encouragement and to Mr. Richard Sadler for coping amicably with my dilatory habits. I hope that the book will not only interest and help students who work mostly on their own, but also that it will give practising and intending mathematics teachers some guidance on a methodical approach to problem-solving.

University College of North Wales S. MOSES
Bangor.
1973

1. Abilities

We are often told that 'mathematics consists in problem-solving'. This is a misrepresentation of the whole of mathematics as a single aspect of it, because there are two distinct sides to the learning of mathematics—the *acquisition of knowledge* and the *testing of one's understanding* of that knowledge. A student usually acquires his store of suitably structured knowledge by instruction. The modern trend in instruction in school mathematics tends to overemphasize the structural aspect of mathematics and, as a consequence, students are often in difficulties when confronted with problems or questions which do not conform to well-known patterns. Yet one purpose of applying our energy to problem-solving is to identify areas of knowledge in which we are not adequately equipped. This automatically suggests that we either revise such work or else make a further study of that particular section of mathematics.

There is no royal road to problem-solving. It will help, however, if we first discuss some of the abilities needed, and later carry on to a study of some general strategies and tactics used to solve problems. The dividing line between abilities needed, tactics to apply, and over-all strategy is not a clear one, and we will sometimes be discussing the same topic under each of these headings. What, however, is quite clear is the vital necessity for a student to have some definite motivation if he is to suceed in solving unusual problems. The most useful personal qualities that give a definite motivation are the *desire for mastery* and a *determination to suceed.*

The desire for mastery is the desire to be able to use our

acquired knowledge in practical contexts, and such desire is much easier to display when the problem has some definite interest to the student. Sometimes the desire for mastery is just plain curiosity.

A self-educated builder once called at my house to ask me to tell him how to estimate heights of objects at sea. He said:

> 'While my men were loading sand at West Shore, I looked towards Puffin Island and realized it appeared only a foot or so above sea-level. Is there any simple formula for allowing for the earth's curvature when estimating the height of a distant object?'

He was delighted when able to estimate such heights with reasonable accuracy. Such curiosity is rare, satisfying, and gives as much pleasure to the tutor as to the inquirer.

The determination to succeed fluctuates with hope and with little successes. It is hard to persevere when we cannot see any way to progress with a solution, but it is easy to carry on when some small advance appears to have been made. Will-power as well as intellect plays an important part in problem-solving. Perhaps self-confidence is the strongest asset in our determination to succeed, but a willingness to ask for help should not be regarded as a lack of self-confidence. In general a student will derive more benefit by sticking at a problem than by asking someone else to do it for him, and so requests for help should not be abused. We should limit requests for help to the obtaining of some useful advice to enable us to proceed further with our problem or to show us why the method we have used has not succeeded. Our method may be a correct one for our particular problem, but sometimes we apply it incorrectly, and sometimes we get bogged down in its application.

Students up to O-level, if properly instructed, are expected to exhibit certain abilities as a result of their instruction. Such abilities could be grouped into the broad categories of *knowledge*, *comprehension*, and *application*. At the present time the great majority of problems likely to be set at O-level are straightforward problems whose solution

requires only a simple understanding of the question and an immediate application of a well-known standard method or technique.

Knowledge

(*a*) The ability to recall facts, nomenclatures, laws, and theories.

(*b*) Suitable experience with practical techniques and straightforward calculations.

Comprehension

(*a*) The ability to translate data from one form to another.

(*b*) The ability to interpret or deduce the significance of data.

(*c*) The ability to choose a suitable method where more than one operation is possible.

Application

(*a*) The ability to apply knowledge and experience to new situations where the method of solution is neither given nor implied.

(*b*) The ability to apply knowledge and experience to old situations presented in a novel manner (such as a new context).

Here are some elementary examples to illustrate the above.

EXAMPLE 1

(*a*) Factorize x^3-6x-9.

The student at this level will know the principle that, if $f(x)$ is a polynomial such that $f(a) \equiv 0$, then $f(x)$ is exactly divisible by $(x-a)$. He soon finds that

$$f(3) = 27-18-9 \equiv 0$$

$\Rightarrow$ † $(x-3)$ is a factor, so

$$x^3-6x-9 = (x-3)(x^2+3x+3) \qquad \text{(on dividing).}$$

† Read the symbol '$\Rightarrow$' as 'implies'.

(*b*) Suppose the question had been:

Factorize $8x^3-12x-9$.

The first time a student at O-level meets this type of example, he probably fails to solve it, although the solution simply needs the application of the same principle. He has, however, to recognize that by rearranging the polynomial in the form

$$8x^3-12x-9 = 8(x^3-\tfrac{3}{2}x-\tfrac{9}{8}),$$

he has to find a suitable value of a that divides exactly into the number $-\frac{9}{8}$. On trying various factors he will eventually find that $f(\frac{3}{2}) \equiv 0$. This, at first sight, may cause him to divide the polynomial by $(x-\frac{3}{2})$, whereas he later discovers that $(2x-3)$ is more convenient than $(x-\frac{3}{2})$. Thus

$$8x^3-12x-9 \equiv (2x-3)(4x^2+6x+3).$$

Example 1(*a*) is a routine example requiring only 'knowledge' for its solution; Example 1(*b*) requires 'knowledge' and also 'application'.

Queries

(i) In solving 1(*a*) we would try $x = 1$ and then $x = 3$. Why is it unnecessary to try $x = 2$?

(ii) Can you see a simple mapping that changes 1(*b*) into 1(*a*) and so give a simple solution of 1(*b*)?

Example 2

Show that the points $A(1,1)$, $B(4,5)$, and $C(5,-2)$ form three vertices of a square and find the coordinates of the fourth vertex.

We first draw a figure (figure 1.1) in case it helps us to find an easy solution. It certainly won't do any harm. We think of the properties of a square and select the most useful ones to fit our particular data. Clearly $AB = AC$ and

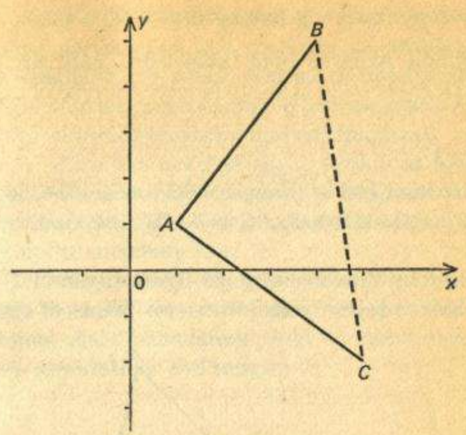


Fig. 1.1

a right angle at A is necessary and sufficient. We know the formula for distance d between (x_1,y_1) and (x_2,y_2):

$$d^2 = (x_2-x_1)^2+(y^2-y_1)^2.$$
$$AB^2 = 3^2+4^2 = 25, \quad AC^2 = 4^2+3^2 = 25,$$
$$BC^2 = 1^2+7^2 = 50$$
$$\Rightarrow AB = AC \text{ and } A \text{ is a right angle (Pythagoras)}$$
$$\Rightarrow A, B, C \text{ are three vertices of a square.}$$

We now try to use some property of the square that will obtain the coordinates of D without giving too much calculation. We should soon remember that the diagonals bisect one another. This gives our result immediately since we know that

$$\left(\frac{x_1+x_2}{2}, \frac{y_1+y_2}{2}\right)$$

is the mid-point of line from (x_1,y_1) to (x_2,y_2). So, mid-point of

$$BC = \left(\frac{9}{2}, \frac{3}{2}\right) \equiv \text{mid-point of } AD = \left(\frac{x+1}{2}, \frac{y+1}{2}\right),$$

hence $x = 8, y = 2.$

This example required both 'knowledge' and 'comprehension'.

EXAMPLE 3

A circle inscribed in triangle ABC touches the side BC at P. If the lengths $AB = 8$, $BC = 7$, $CA = 6$, find the length of BP.

We start by drawing a *rough* figure (figure 1.2) and it seems obvious to make the other two points of contact Q

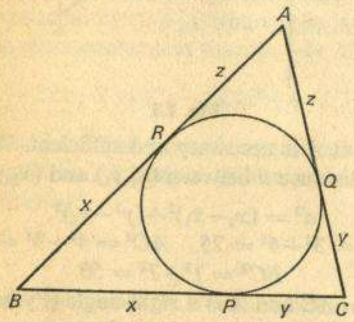


Fig. 1.2

and R. We know that two tangents to a circle from the same point are equal in length, and we can see that the sides of the triangle are tangents to the circle.

How are we going to connect up these two facts? We mark the equal lengths and may think of writing them as lengths x, y, and z. Now we realize that we can *change the geometry problem into an algebra problem.*

Perimeter of triangle $= 2x+2y+2z = 8+7+6,$

giving $$x+y+z = 10\tfrac{1}{2}.$$

So, $$x = (x+y+z)-(y+z) = 10\tfrac{1}{2}-6 = 4\tfrac{1}{2},$$

hence length of BP is $4\tfrac{1}{2}$.

This example illustrates 'knowledge' plus 'application'. As mentioned earlier, the vast majority of problems up to O-level standard are either routine or reasonably simple variations of routine problems. When we reach A-level or university entrance mathematics we frequently encounter problems which are not just slight variations of standard problems. We now need to consider in greater detail the abilities required for solving mathematical problems in general up to university level.

Again these abilities are grouped in broad categories, although there will be a frequent overlap between 'comprehension' and 'formulation', and an occasional overlap between 'application' and insight.

Comprehension: The ability to interpret and deduce the significance of data.

Formulation: The ability to find a suitable mathematical model.

Knowledge: The ability to recall facts, theorems, nomenclatures, and techniques.

Manipulation: The ability to perform operations on symbols and to carry out calculations, sometimes lengthy and not straightforward.

Application: The ability to apply knowledge and experience to new situations or to old ones in a new context.

Insight: The ability to apply one's imagination to an unusual situation, to think of algorithms or analogues when no method of solution is evident.

It must be stressed that the close interweaving of the simple logical abilities inherent in comprehension and formulation with the creative abilities needed for application and insight make the separation of the various aspects of

mathematical abilities very difficult. In general the most important mathematical abilities needed for problem-solving are *self-involvement* and a combination of *application* and *intuition* followed by *logical deduction*. Before giving illustrative examples we must consider further the ability to manipulate.

In the 'good old days' manipulation in schools became an end in itself and so became very boring to the intelligent student. Its biggest vice was the neglect to teach understanding of the process involved, which meant that there was no transfer from a particular process to any new situation where the same underlying principle was involved. For example, we learned at the age of about 12 the method for finding the highest common factor of two very large numbers.

EXAMPLE 4

Find the H.C.F. of 2,183 and 4,189.

```
2,183)4,189(1
      2,183
      -----
      2,006)2,183(1
            2,006
            -----
              177)2,006(11
                  1,947
                  -----
                     59)177(3
                        177
                        ---
                        ...
```

Hence H.C.F. = 59.

The simple principle underlying the process was not explained—'any common divisor of two numbers A and B also divides into the number $A-\lambda B$ for all integers λ'. Had we understood this principle, we would have quickly solved

other examples encountered in algebra much later on, such as:—

EXAMPLE 5

Find all the values of k for which the expressions x^2+kx+6 and $2x^2+kx-3$ have a common factor.

$$A = 2x^2+kx-3, \quad B = x^2+kx+6, \quad \lambda = 1,$$

hence any common factor divides into

$$A-B = x^2-9 = (x-3)(x+3),$$

and hence the only possible common factors are $(x-3)$ and $(x+3)$.

When $(x+3)$ is a factor,

$$\left.\begin{aligned} x^2+kx+6 &= (x+3)(x+2) \Rightarrow k = 5. \\ 2x^2+kx-3 &= (x+3)(2x-1) \Rightarrow k = 5. \end{aligned}\right\}$$

Similarly when $(x-3)$ is a factor, $k = -5$ for each expression

$$\Rightarrow k = 5 \quad \text{or} \quad -5.$$

Yet we did learn to perform certain essential manipulations and computations quickly and without undue mental effort. A facility in manipulation should be cultivated by every student, because it gives a great boost to self-confidence. Yet we must always keep in mind the intuitive meaning of a problem and judge whether our answers are reasonable; otherwise it is very easy to get bogged down in a morass of meaningless manipulations. In other words, a facility in manipulation without understanding of the principle behind that manipulation does not lead to any real mathematical progress, but understanding of such principles without the ability to apply it with reasonable ease is a definite handicap to success in problem-solving.

I now give some examples to illustrate some of the abilities mentioned above.

Example 6

Find the sum of the numbers in the first n brackets of the series

$$(1)+(2+3)+(4+5+6)+(7+8+9+10)+\ldots.$$

Comprehension

When brackets are removed we have a simple arithmetic progression $1+2+3+4+5+\ldots.$

Formulation

This is a standard mathematical model where we need to find the number of terms N.

Knowledge

The sum to N terms is $S_N = \frac{1}{2}N(N+1)$.

Application

The first bracket has one number, the second 2, the third 3, and so on. Clearly the last bracket will have n numbers.

Hence $\quad N = 1+2+3+\ldots+n = \dfrac{n}{2}(n+1).$

Manipulation

$$S_N = \frac{N(N+1)}{2} \text{ where } N \to \frac{n}{2}(n+1)$$

$$= \frac{n(n+1)}{4}\left[\frac{n(n+1)}{2}+1\right]$$

$$= n(n+1)(n^2+n+2)/8.$$

Example 7

Prove that $3^{2n}-1$ is divisible by 8 for all positive integers n.

Comprehension

$3^{2n}-1 = 8N$ is required.

Knowledge

Some students would think immediately of a proof by mathematical induction. Others may notice it is a difference of two squares $\Rightarrow 3^{2n}-1 = (3^n-1)(3^n+1)$.

Manipulation

(*a*) When $n = 1$, $3^{2n}-1 = 8$ is divisible by 8.

Suppose $f(n) = 3^{2n}-1$ is divisible by 8 for some value $n = k$, $\Rightarrow 3^{2k}-1 = 8N$, where N is an integer.

$$f(k+1) = 3^{2k+2}-1 = 9(8N+1)-1 = 8(9N+1).$$

Hence, if $f(k)$ is divisible by 8, so is $f(k+1)$.

Since result is valid for $n = 1$, it is now established for all positive n by mathematical induction.

(*b*) $$3^{2n}-1 = (3^n+1)(3^n-1).$$

Application

We apply well-known facts to a new situation—the facts that, if p is an odd integer, then p^n is an odd integer for all $n \in \mathbf{Z}^+$ and $p\pm1$ must be even integers. Also that, for all positive integers p the numbers $p\pm1$ are consecutive integers.

Solution

Let $p = 3^n$.

$$3^{2n}-1 = (3^n-1)(3^n+1) = (p-1)(p+1)$$

= product of consecutive even integers and so one is divisible by 2 and the other by 4,

hence $3^{2n}-1$ is divisible by 8.

When this question was set in an A-level examination, the great majority of the correct solutions used mathematical induction. Many candidates did factorize but could proceed no further. The idea that $3^n \pm 1$ gives a pair of even integers proved hard to see, and the further realization that these even integers are consecutive was beyond all but two candidates out of 1,500.

EXAMPLE 8

Here is an amusing problem from an old problem journal. A cynic once said that versatility was a thousand times more profitable than veracity. Demonstrate this mathematically.

Comprehension

A thousand times more $\equiv$ 1,001 times as big.

Formulation

$X = 1{,}001\,Y.$

Knowledge

What happens in multiplication by 1,001:

```
      VERACITY
   VERACITY
   -----------
   VERSATILITY
```

Manipulation

We must find a set of numbers 0,1,2, . . .,9 to map one-one on the letters A, C, E, I, L, R, S, T, V, and Y. If we can find only one such set, then the model can be said to be true mathematically. The simple but tedious method of trial and error shows that the mapping

$0 \to$ I	$5 \to$ S
$1 \to$ A	$6 \to$ L
$2 \to$ Y	$7 \to$ E
$3 \to$ V	$8 \to$ R
$4 \to$ C	$9 \to$ T

does satisfy the addition sum

```
 VERAC
 ACITY
 -----
 SATIL.
```

To show this mapping is unique is more difficult, but it can be shown by the method of exhaustion.

The cynic's statement has been demonstrated mathematically.

The small collection of exercises given at the end of the book is not arranged in order of difficulty. Some of the exercises are purely tests of manipulative ability, some are tests of application, and some require a little insight. Starred questions should be regarded as more difficult either because the knowledge needed may be above that of A-level mathematics or because some definite insight is required to find a suitable method of attacking the problem.

2. Strategies

The majority of problems at A-level are 'routine' problems which are solved either by substituting special data into a general result already proved or by following, step by step, some well-known method. Routine problems are, of course, essential if we are to acquire facility in manipulation, but, as no originality is required, skill in the mechanical performance of routine mathematical operations does not help very much in finding procedures useful for solving problems in general. To make progress in general problem-solving we need to formulate an over-all strategy for tackling problems in general. Usually, we begin by expecting to find many types of strategy to fit the many distinct types of problems. We are, however, muddling 'strategy' with 'tactics'. A general strategy should apply to problems, both mathematical and non-mathematical. What specific problems require is usually a distinct method of approach and, perhaps, a different method of proof. We will discuss methods of proof in a later chapter, but here we will study the beginnings of a general strategy.

Problems in general divide into two types—the 'to find' type and the 'to prove' type. The latter formed the major part of the traditional geometry course in grammar schools, and we will start by considering the following plan for tackling the 'to prove' type of problems in geometry which I used to give to my pupils in an upper school.

A. THE QUESTION

1. Read it carefully several times until you understand it fully.
2. Underline any statement or phrase which gives or implies some definite fact about the figure.

B. THE FIGURE

1. Draw a rough figure on scrap paper to see how it can be drawn to best advantage in your book.
2. Draw the most convenient figure that fits *all* the data.
3. Make sure you have not drawn a *special* figure.
4. Does the result *look right* in your figure?

C. THE DATA

1. Mark all data in *ink* in your figure. If there is some property you cannot so mark, write it alongside your figure.
2. Remember that words like 'rectangle', 'cyclic', etc., imply properties of the figure that *may be needed in the proof.*

D. DEDUCTIONS

1. Mark any obvious deductions in *pencil* in your figure and try to obtain the required result from these deductions.
2. If the problem is in two parts, the first part is usually needed to prove the second part.
3. If you cannot prove the first part, assume it to be true and use it to prove the second part.

E. THE PROOF

1. If successful, remember to give reasons for all statements *not* marked in *ink* in your figure.
2. If you can prove part of the question, do so.

F. FAILURE

Ask yourself the following questions:

1. Have I used *all the data* in the question?
2. Is an *auxiliary line* needed in the figure?
3. Have I solved a similar problem before? If so, look up your proof in case the same ideas are useful here.
4. Can I *restate the result* in a different way? The new way may be easier to solve.

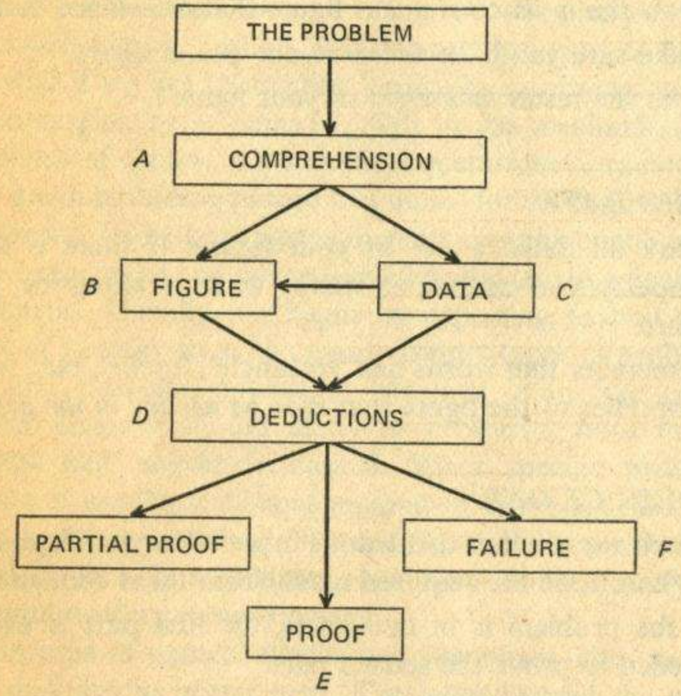


Fig. 2.1

5. Can I *work backwards*? (i.e. mark the result in your figure and see what follows from it).
6. Have I remembered the word 'implies' in A2?
7. Does an *accurately* drawn figure help me?
8. Can I prove a *special case* of the problem?

If we try to abstract a general strategy from this advice for solving geometry problems, we would get a schematic representation something like figure 2.1.

Now we wish to make some transfer of the strategy for solving geometry problems to a strategy for solving problems in general. A reasonable first attempt at such abstraction might be as shown in figure 2.2.

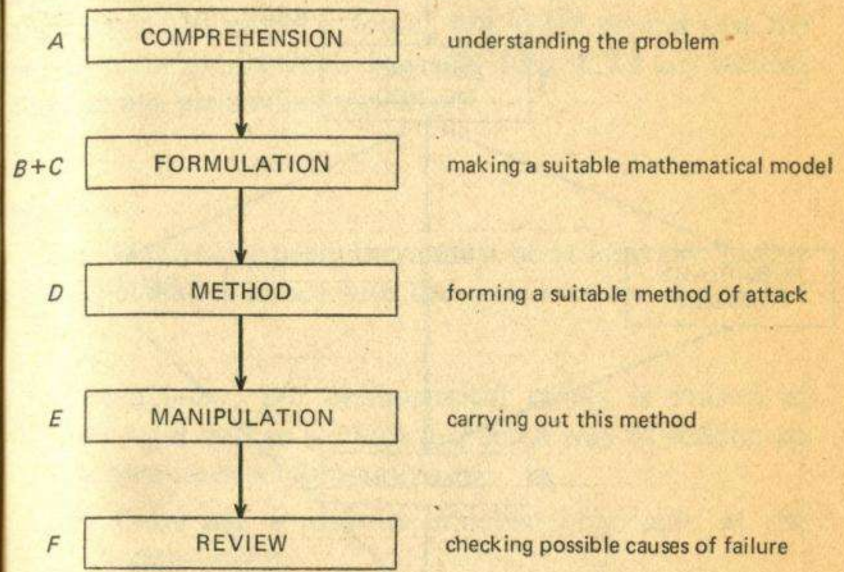


Fig. 2.2

Clearly a crude abstraction such as this requires more precision if we wish to consider it as a general over-all strategy for problem-solving. Again, there will be a difference in tactics between the 'to prove' type of problem and the 'to find' type. In the former the required result is given, whereas in the latter we should obviously check at all possible stages whether our results are physically reasonable. Also it will be useful to consider whether the result obtained or the method adopted can be applied to 'related' problems. This last point will be discussed in the last chapter under 'problems and their extensions'. Here we will

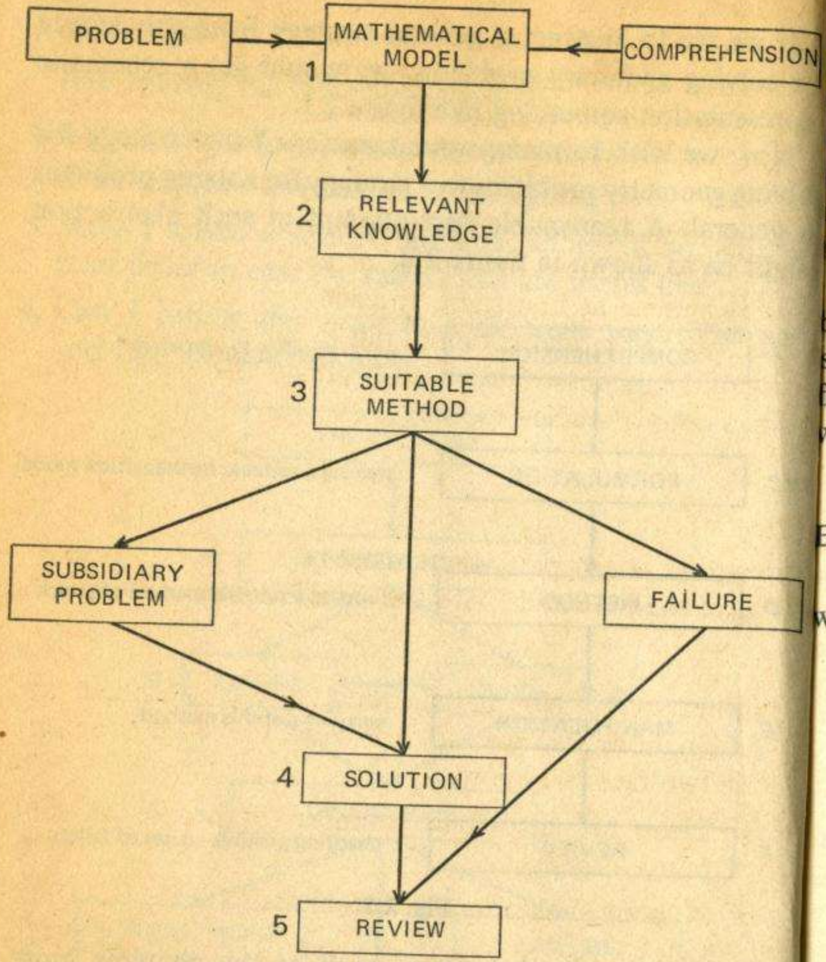


Fig. 2.3

study a more sophisticated model of the above strategy for general problem-solving (figure 2.3).

We will now consider more fully the various stages in the above strategy together with some illustrated examples.

1. MATHEMATICAL MODEL

This is the combining of the problem and its comprehension into the formulation of a mathematical model to

which we can apply known rules and operations in the approximate mathematical context. Sometimes the model is self-evident in the problem as stated, sometimes a little thought will be needed to choose a suitable model, and occasionally deep study may be required to provide a model clear enough for further progress.

Here are two illustrative problems which hardly seem to be mathematical at first sight. In each case the solution is simple once the model is found, but in the second case the finding of the model would be really difficult for any student who has not yet studied topology.

EXAMPLE 1

Show that at any party there must be at least two people who have shaken hands with the same number of people.

Investigation: No mathematical model is evident at first sight and so we look for some way of linking up the two facts we know:

(i) There are a definite number of people at the party—say n,

(ii) Nobody shakes hands with himself (while sober!).

Suppose we represent the set of people by $P = \{p_1, p_2, \ldots, p_n\}$ and the set of their possible number of handshakes by $S = \{0, 1, 2, \ldots, n-1\}$. If we can show the mapping $P \to S$ is not one-one, we know that at least two elements of P map on to the same element in S, which is what we have to prove.

Mathematical model: $P \to S$ is a mapping.

Relevant knowledge: P and S each contains n elements.

Suitable method: Show the mapping is not one-one.

Solution: A proof by contradiction suggests itsel immediately.

Suppose the mapping P → S is one-one.

Let p_r be the person who shakes hands 0 times an p_s the person who shakes hands $(n-1)$ times. Thi gives a contradiction because p_s has shaken hands wit everybody else and so there cannot exist a person p who has shaken hands with nobody.

Hence mapping P → S is not one-one.

Hence at least two elements of P map onto one of th elements in S.

EXAMPLE 2

We are given a 3 × 3 chessboard with white knights (KT) i the two top corners and black knights (kt) in the two botto corners. We have to interchange the white knights with th black knights in the least number of moves.

Investigation

A tedious solution by trial and error is possible, bu how do we know when all the possible combinatio of moves have been exhausted? So we try to find mathematical model. It seems reasonable to numbe the squares and then to write down some of the startin moves.

KT 1 can move to square 6 or 8.

KT 3 can move to square 4 or 8 (see figure 2.4).

We may think sooner or later of drawing lines in ou diagram to indicate all the possible starting moves, fou of which are marked in the top right of figure 2.4 Would we have the insight to think of these lines a strings and the numbers as buttons? Suppose, howeve we stretch out the square board until its boundar

becomes a circle, giving us the third figure. If we again draw in all the possible moves, starting with move 1 → 6, we soon realize that we end up on square 1, where we had started. Now it is easy to see that if we think of these lines as strings, we have only to join up

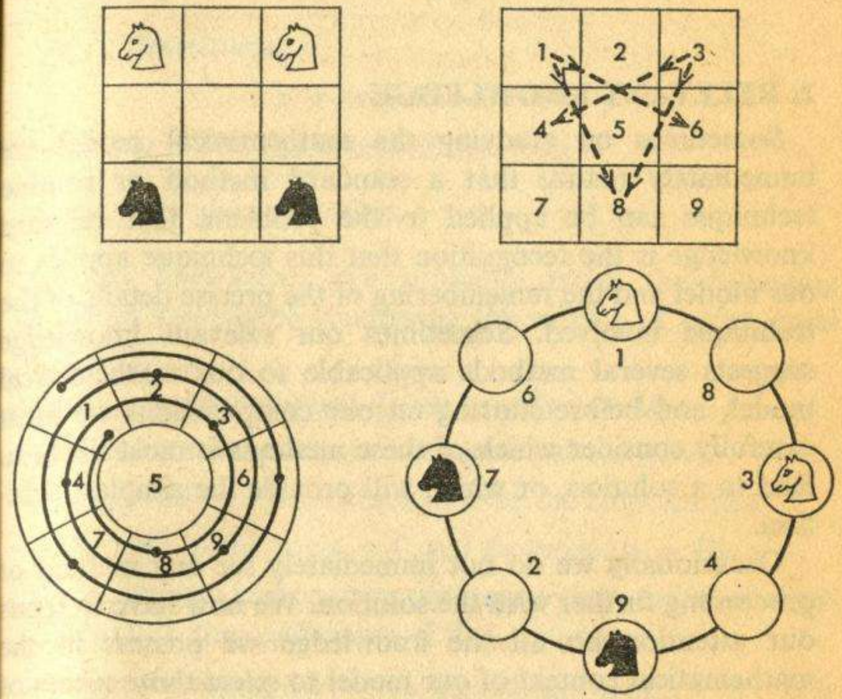


Fig. 2.4

the two ends in square 1 to get a circle as shown in the fourth figure. The squares have become buttons and the moves become strings. Here now is our mathematical model.

Solution

Opening out the strings to make a circle does not change the topological structure of the squares 1 to 9 and their connection with the knight's move. Hence we simply have to keep moving the knights round the circle in one direction until they are interchanged.

This will give us 16 moves. Remember that in our model when one knight moves, all the four knights must move. If we keep moving clockwise on the board the sequence is $1 \to 6$, $7 \to 2$, $9 \to 4$, $3 \to 8$, $6 \to 7$, $2 \to 9$, $4 \to 3$, $8 \to 1$, $7 \to 2$, $9 \to 4$, $3 \to 8$, $1 \to 6$, $2 \to 9$, $4 \to 3$, $8 \to 1$, $6 \to 7$.

2. RELEVANT KNOWLEDGE

Sometimes on studying the mathematical model we immediately realize that a standard method or routine technique can be applied to the problem. Our relevant knowledge is the recognition that this technique applies to our model and the remembering of the precise details of the technique involved. Sometimes our relevant knowledge suggests several methods applicable to our mathematical model, and before starting on our computations we must carefully consider which of these methods is most likely to lead to a solution, or which will provide the simplest solution.

Occasionally we do not immediately see any method of proceeding further with the solution. We now have to focus our attention on all the knowledge we possess in the mathematical context of our model to select those pieces of knowledge which seem to have some connection with our particular model. Now we concentrate on these relevant pieces of knowledge—facts, laws, theories, and so on—and try to find any which seem to have a very direct bearing on our problem. Paradoxically, too much knowledge in a particular mathematical context may handicap us in selecting the most appropriate relevant knowledge in the sense that we 'fail to see the wood for the trees'.

Lastly, our solution may sometimes depend on our starting point. Student *A* has perhaps covered a wider syllabus in a particular mathematical context than student *B* has. As a result student *A* may immediately recall a special piece of

relevant knowledge which bears very directly on a particular model and which leads to a short straightforward solution. Student *B* may find this problem far more difficult because he does not possess that special piece of relevant knowledge.

EXAMPLE 3

Solve the equation

$$x^4+3x^2+6x+10 = 0,$$

given that $x = 1+2i$ is one of the roots.

Model: The given equation.

Relevant knowledge: We probably recall the following pieces of knowledge concerning the solving of polynomial equations.

1. Any integer root must divide exactly into 10.
2. Complex roots occur in conjugate pairs $a \pm ib$.
3. There will be 4 linear factors over the complex field.
4. The sum of the roots $= 0$ and the product $= 10$.
5. There are no real positive roots since there are no changes of sign in the coefficients of $f(x)$, and so on.

On considering the further datum that $x = 1+2i$ is a root, we soon realize that facts 2 and 3 bear very directly in our particular model and will lead to a suitable method.

Suitable method: $f(x) = 0$.

$x = 1+2i$ and $x = 1-2i$ are both roots (fact 2),

hence $(x-1-2i)(x-1+2i)$ are two of the factors of $f(x)$,

hence (x^2-2x+5) divides exactly into $f(x)$, so

$$f(x) = (x^2-2x+5)(x^2+2x+2)$$

$$\Rightarrow f(x) = 0 \text{ has roots } x = 1 \pm 2i,\ -1 \pm i.$$

Example 4

Solve the difference equation

$$u_{n+2} - 5u_{n+1} + 6u_n = 0 \text{ for all } n \in \mathbf{Z}^+.$$

Model: Equation as given.

Comprehension: We want the general function of n to satisfy the above equation for all $n \in \mathbf{Z}^+$.

Relevant knowledge: Student A may have no knowledge of difference equation theory and so has to rely on insight and manipulative skill to obtain a solution.

Solution: A first flash of insight comes with a realization that the substitution $v_n = u_{n+1} - 2u_n$ or $v_n = u_{n+1} - 3u_n$ will reduce the equation to first order:

$$(u_{n+2} - 2u_{n+1}) - 3(u_{n+1} - 2u_n) = 0,$$

hence $v_{n+1} - 3v_n = 0$ where $v_n = u_{n+1} - 2u_n$ for all $n \in \mathbf{Z}^+$

$$\Rightarrow v_{n+1} = 3v_n = 3^2 v_{n-1} = \ldots = 3^n v_1$$

hence $v_n = a \,.\, 3^n$ where a is an arbitrary constant. So,

$$u_{n+1} - 2u_n = a \,.\, 3^n.$$

A second flash of insight comes with the realization that the other substitution will similarly give

$$u_{n+1} - 3u_n = b \,.\, 2^n,$$

hence $u_n = a \,.\, 3^n + b \,.\, 2^n$, where a and b are arbitrary.

Student B may have read a little difference equation theory and will recall the relevant theories—'If k and l are constants, $l \neq 0$, then the difference equation

$$u_{n+2} + ku_{n+1} + lu_n = 0$$

is satisfied for all $n \in \mathbf{Z}^+$ by the general solution

$$u_n = A\alpha^n + B\beta^n \ (A, B \text{ constants}),$$

where α, β are the distinct roots of the quadratic equation

$$x^2 + kx + l = 0.\text{'}$$

Solution: $x^2 - 5x + 6 = 0 = (x-2)(x-3)$,
hence α, β take the values 2, 3

$$\Rightarrow \text{general solution is } u_n = A \,.\, 2^n + B \,.\, 3^n.$$

Notice that further relevant knowledge in Example 3 did not lead to a simpler solution, but in Example 4 it made the problem a routine exercise. The latter emphasizes the point that the solution will depend to a great extent on the relevant knowledge available for a particular problem by a particular student. Sometimes too much knowledge can be a handicap because it focuses our attention on certain procedures suggested by that knowledge and causes us to overlook simple solutions.

3. SUITABLE METHOD

Here the approach to the 'to prove' type of problem may differ from that to the 'to find' type. In the former the problem is given in the form of a proposition and we concentrate all the relevant knowledge on finding an appropriate method for proving that particular proposition. We do not think of the various methods of proof as being different, but simply that each represents a distinct tactical approach to the same end, the proving of the proposition. At first we are satisfied with any proof, no matter how lengthy, that leads to the desired result, but as our experience grows we learn to avoid tedious proofs and to look for short cuts. Sometimes the only difference between a long tedious proof and a short elegant one is the level at

which we encounter the problem. A little piece of more advanced knowledge may revolutionize our method of proof of a particular proposition, as shown in the last example. Methods of proof will be discussed in detail in Chapter 4 and so we will omit illustrative examples here.

In the 'to find' type we have two distinct approaches—one for problems where no conjecture can be found and another for those where a conjecture is possible. In the former type we focus all our relevant knowledge of routine methods and techniques to choose a possible one for our particular problem. A routine problem can always be solved by the appropriate standard technique, but a closer examination of the data of a particular problem may suggest a variation of the standard technique or a special method that will provide an easier solution. Of course a solution by a long method is always acceptable, but there is personal satisfaction in the finding of an individual proof that avoids complicated workings and leads to a neater and shorter solution. In the type where we first obtain a conjecture, we are really changing such problems to the first type—the 'to prove' type. A conjecture is usually found by applying the various approaches of the inductive process—observation, analogy, and insight. These will be discussed fully in the next chapter, and so the following examples illustrate some of the earlier points.

EXAMPLE 5

(*a*) Evaluate $$\int_0^1 \frac{x\,dx}{1+x^4}.$$

Investigation

Since the integrand is a rational function of x, we know that we can evaluate the integral by finding the partial fraction equivalent of the integrand and then

evaluating each term separately. However, we require simple insight to factorize the denominator:

$$x^4+1 \equiv (x^2+1)^2-2x^2 \equiv (x^2-\sqrt{2}x+1)(x^2+\sqrt{2}x+1).$$

Thus we examine the problem more closely in case some other method will lead to an easier solution. We soon realize that the substitution $y = x^2$ will lead to a simpler integral:

$$y = x^2 \Rightarrow dy = 2x\,dx \Rightarrow \int_0^1 \frac{x\,dx}{1+x^4} = \tfrac{1}{2}\int_0^1 \frac{dy}{1+y^2},$$

hence $$\int_0^1 \frac{x\,dx}{1+x^4} = \tfrac{1}{2}[\tan^{-1}y]_0^1 = \tfrac{1}{8}\pi.$$

(*b*) Find the coefficient of x^{10} in the expression of $(1+x+x^2)^{-1}$ in ascending powers of x.

Investigation

We know the given expression can be expanded binomially as $(1+z)^{-1}$, where we replace z by $(x+x^2)$, but we will need to evaluate all terms up to that involving $(x+x^2)^{10}$. Suppose we examine the given expression more closely:

$$\frac{1}{1+x+x^2} = \frac{1-x}{1-x^3} = (1-x)(1-x^3)^{-1}$$

$$(1+x+x^2)^{-1} = (1-x)(1+x^3+x^6+x^9+x^{12}+\ldots),$$

hence the coefficient of $x^{10} = -1$.

4. FAILURE

(*a*) Failure to find a possible method

The commonest cause of our failing to find a possible method is failure to understand the problem fully. We fail to take into account all the essential ideas involved in a problem. It may be necessary to reconsider each part of the data in detail, even going back to the definition of the terms.

This may help to introduce new ideas which may lead to a more accessible problem. Somehow we must reconsider *all* the ideas suggested by the *whole* data to obtain a possible method of investigating the problem further. In other words:

1. HAVE WE USED ALL THE DATA?

EXAMPLE 6

Solve the equation

$$x^4-3x^3+2x^2-3x+1=0.$$

Model: Equation as given.

Comprehension: To find the four roots α, β, γ, δ (real or complex).

Relevant knowledge: We would probably think about:

(i) The equation can be written $(x-\alpha)(x-\beta)(x-\gamma)(x-\delta) = 0$.

(ii) Any integer root must be a factor of 1.

(iii) Complex roots must occur in conjugate pairs and so give rise to a quadratic factor with real coefficient, because if $\alpha = a+ib$,

$$(x-\alpha)(x-\bar{\alpha}) \equiv x^2-2ax+a^2+b^2.$$

Suitable method: We see that $x = \pm 1$ do not fit the equation and so there are no integer roots. We probably conjecture that there will be two quadratic factors, each giving a pair of conjugate roots. However, we do not know any standard method of factorizing quadrics in general. We know from experience that the use of the theory of equations will not help us, because in the application we eventually end up with an equation of the same form as the given equation. Except for guessing we appear to be stuck!

Have we used all the data? Have the data any special features? When we reconsider the equation in detail, we notice that the coefficients are exactly the same in each direction. This might suggest to us that we pair off the terms with the same coefficients, giving us:

$$(x^4+1)-3x(x^2+1)-2x^2 = 0.$$ wrong

From experience we know that equations in one variable are often solved by considering a subsidiary equation of lower degree. Hence we should sooner or later think of the substitution $y = x^2+1$.

Hence $\quad (y^2-2x^2)-3xy-2x^2 = 0.$

$$y^2-3xy-4x^2 = 0 = (y-4x)(y+x).$$

Hence $\quad y = 4x \Rightarrow x^2+1 = 4x \Rightarrow x = 2\pm i\sqrt{3}.$

$$y = -x \Rightarrow x^2+1 = -x \Rightarrow x = -\tfrac{1}{2}\pm i\frac{\sqrt{3}}{2}.$$

Note: Equations of this type are called 'reciprocal' equations because the mapping $x \to (1/x)$ does not change the form of the equation. The first step in the solution is usually written as

$$\left(x^2+\frac{1}{x^2}\right)-3\left(x+\frac{1}{x}\right)-2 = 0.$$

The substitution $y = x+(1/x)$ is now very obvious, but it is unlikely that a student solving such an equation for the first time would realize this fact.

Again, it may be possible to prove a part of the problem, but it is necessary that our new unknown should be more accessible than the original one. Also our new unknown must be useful in the sense that it will help us towards finding the original unknown. Such ideas lead to 'subsidiary problems' or 'auxiliary problems'. The former refer to those ideas where the new unknown is more accessible and has an

obvious definite bearing on the finding of the original unknown, whereas the latter have only a vague connection with the original problem. When stuck we should certainly consider such auxiliary problems because further study may clarify the connection between our new result and the original one. After all, even vague suggestions towards a possible method of attacking the original problem are better than none at all.

Further, we sometimes encounter problems in which some apparently extraneous data is given. Usually the use of this extra data leads to a subsidiary problem more accessible than the original one. The setter of the problem may be doing this deliberately because he feels that without this extra help the problem will be too difficult for the student to solve. Such extraneous data is often in the form of a suggestion towards a certain course of action rather than a further definite fact about the mathematical model. In other words:

2. CAN WE FIND A SUBSIDIARY PROBLEM?

Again it may happen that we have previously come across and solved a similar problem. We should look up our solution of this problem and study it carefully to see whether it is in any way related to the problem in hand. The more closely related it is, the more likely that the same method of attack, perhaps varied slightly, may apply to the problem. In thinking of relevant knowledge we tend to concentrate on previously proved propositions, whereas a proper mobilization of our previously acquired knowledge should include the problem we have solved as well as the propositions we have studied. In other words:

3. HAVE WE SOLVED A SIMILAR PROBLEM BEFORE?

EXAMPLE 7

Find the sum for positive integer n of the series

$$1+2\binom{n}{1}+3\binom{n}{2}+4\binom{n}{4}+\ldots+(n+1)\binom{n}{n}.$$

We could use specialization with $n=1, 2, 3, 4$ and try to make a conjecture for the possible result. But we may remember that we have solved a similar type of series before in the form

$$1+\binom{n}{1}+\binom{n}{2}+\ldots+\binom{n}{n}=2^n.$$

We will see whether the result or the method of proof will help us here.

Proof.

$$S(x)=1+\binom{n}{1}x+\binom{n}{2}x^2+\binom{n}{3}x^3+\ldots+\binom{n}{n}x^n$$
$$=(1+x)^n,$$

hence $S(1)=1+\binom{n}{1}+\binom{n}{2}+\ldots+\binom{n}{n}=2^n.$

It should not take us long to see what is needed to apply this method to our particular problem.

$$F(x)=x+\binom{n}{1}x^2+\binom{n}{2}x^3+\ldots+\binom{n}{n}x^{n+1}=x(1+x)^n.$$

Differentiate:

$$F'(x)=1+2\binom{n}{1}x+3\binom{n}{2}x^2+\ldots+(n+1)\binom{n}{n}x^n$$
$$=(1+x)^n+nx(1+x)^{n-1},$$

hence

$$F'(1)=1+2\binom{n}{1}+3\binom{n}{2}+\ldots+(n+1)\binom{n}{n}$$
$$=2^n+n\,.\,2^{n-1},$$

hence our required sum is $(n+2)2^{n-1}$.

Again a problem may prove quite difficult until we look at it in a different light. The simplest way of finding a new way of attacking a problem is to restate the problem in a different way. It may cause us to recall that we have seen the same problem before but in a slightly different form. Sometimes an awkward problem becomes quite simple when restated in a different way. Sometimes the terms used in the problem may be capable of more than one definition, and our reactions to the problem will partly depend on the choice we made from the possible definitions. Consider, for example, two distinct definitions of the *rank of a matrix*:

(*a*) The number of linearly independent rows.

(*b*) The number r such that at least one r-rowed minor is not zero, and all $(r+1)$-rowed minors are zero.

If we have proved these are logically equivalent, certain propositions may be much easier to prove with one definition than with the other. For example, the second definition leads immediately to the theorem that the number of linearly independent columns of a matrix is the same as the number of linearly independent rows.

It is even possible that our problem can be restated in more than one way and, when stuck, it is essential to consider each of these ways. What often happens is that the restatement of the problem brings to mind some previously acquired knowledge that bears directly on our problem. We might possibly have previously recalled this particular relevant knowledge from a fuller understanding of our mathematical model, but the restatement focused our attention on it. In other words,

4. CAN WE RESTATE THE PROBLEM DIFFERENTLY?

EXAMPLE 8

In how many ways can 4 suspects be chosen from an identity parade of 12 persons without choosing any two consecutive persons?

Investigation

We can, of course, solve this problem by the lengthy process of exhaustion of all the possible choices. Other direct methods also seem to give long solutions. Yet the problem becomes very elementary when we restate it as follows:

Restatement

In how many ways can we place 4 ladies into a row of 8 gentlemen so that no two ladies are adjacent?

Solution

Each man can have a lady on either side of him. Eight men will provide 9 spaces in which ladies can be placed,

hence number of ways $= {}^9C_4 = \dfrac{9\,.\,8\,.\,7\,.\,6}{1\,.\,2\,.\,3\,.\,4} = 126.$

Lastly, have we tried to work backwards? Most geometrical textbooks give some explanation of *analysis* and *synthesis*. In analysis of the 'to prove' type of problem we start from what we have to prove, i.e. the conclusion of the proposition. We assume this to be true and try to find an antecedent or previous step from which this conclusion could have been obtained. We then carry on to look for an antecedent of this antecedent, and so on. Eventually we hope to arrive at the data of the problem or some other result admittedly true. The better-known name of such analysis is 'working-backwards'!

The changing of our procedure of analysis, when successful, into a proof is called *synthesis*. We reverse the process starting from the data or some other logically equivalent condition and retrace the steps in our analysis until we finally arrive at the required conclusion. This 'working backwards' procedure is not so fruitful in the 'to find' type

of problem. Where it can be used, we start by assuming an answer X exists. We then try to find another unknown Y which would result from the existence of X. The condition satisfied by Y may not be the same as that satisfied by X but it will be a related condition. We carry on until we reach some unknown Z which is accessible, i.e. which can be ascertained by a known method. If each step is reversible, we simply retrace our steps, starting with Z and ending with the required result X. In other words,

5. HAVE WE TRIED 'WORKING BACKWARDS'?

EXAMPLE 9

Prove that $\sin(\cos x) < \cos(\sin x)$ for $0 \leqslant x \leqslant \frac{1}{2}\pi$.

Analysis: Assume the result is true and work backwards:

$$\sin(\cos x) < \cos(\sin x)$$

is the same as

$$\sin(\cos x) < \sin(\tfrac{1}{2}\pi - \sin x).$$

When $x \in [0, \frac{1}{2}\pi]$, $\cos x$ lies in $[0, 1]$ and $\frac{1}{2}\pi - \sin x$ lies in $[\frac{1}{2}\pi - 1, \frac{1}{2}\pi]$. Thus the arguments on each side of the inequality are acute angles. But $\sin\theta$ is a monotonic increasing function for θ an acute angle,

hence the inequality will be satisfied if $\cos x < \frac{1}{2}\pi - \sin x$,

hence $\cos x + \sin x < \frac{1}{2}\pi$.

This is our *subsidiary* problem; if we can prove this we can retrace our steps and prove the original problem.

Solution: $\cos x + \sin x = \sqrt{2}\cos(x - \frac{1}{4}\pi)$,

hence $|\cos x + \sin x| \leqslant \sqrt{2} \Rightarrow |\cos x + \sin x| < \frac{1}{2}\pi$, since $\frac{1}{2}\pi > \sqrt{2}$.

Hence $\cos x + \sin x < \frac{1}{2}\pi \Rightarrow \cos x < \frac{1}{2}\pi - \sin x$.

When $x \in [0, \frac{1}{2}\pi]$ both sides of the inequality are positive and their values in *radians* will represent acute angles. But $\sin\theta$ is a monotonic increasing function of θ in the interval $[0, \frac{1}{2}\pi]$.

Hence $\sin(\cos x) < \sin(\frac{1}{2}\pi - \sin x)$

$$\Rightarrow \sin(\cos x) < \cos(\sin x).$$

The inequality is true for all real values of x, but the above proof would then need extensions and careful attention to signs, starting with $\sin(\cos x) < \sin(\frac{1}{2}\pi \pm \sin x)$.

(b) Failure to obtain a correct result

Sometimes we are sure that the method adopted to solve a particular problem is a correct one and yet the answer obtained is clearly wrong from physical considerations. There are two points involved here. The first is that we may have made a computational error. We guard against this as far as is possible by checking every step that can be checked. Above all one must always check the final result whenever this is possible. In a general problem the result can often be checked by specialization, especially by using the extreme cases. In other words:

6. HAVE WE CHECKED EACH STEP?

The second point is that the chosen method may not be really appropriate for the problem, or that it needs some variation before it should be applied. We tend to rush into calculations without having a complete understanding of the over-all plan for the problem. Even where such tactics do eventually lead to a solution, it is usually far more lengthy than is necessary.

EXAMPLE 10

Find the length of the chord intercepted on the line $3x+4y = 7$ by the circle $x^2+y^2-2x-4y+1 = 0$.

Solution 1. The headstrong student immediately embarks on the straightforward method of solving the two equations to find the coordinates of the points of intersection.

Substituting $y = (7-3x)/4$ and simplifying carefully gives

$$25x^2-26x-47 = 0$$

Hence $x = \frac{1}{25}(13\pm\sqrt{1{,}344})$. (Collapse of student!) What he has overlooked is that a *straightforward method does not guarantee simple manipulative workings.*

Solution 2. A thoughtful student begins by drawing a rough figure and studying it to see whether a shorter method is suggested by the figure. He remembers that the perpendicular from the centre of a circle bisects the chord and so produces the following simple solution:

Circle is $(x-1)^2+(y-2)^2 = 2^2 \Rightarrow$ centre $(1, 2)$ and radius 2.

Length of perpendicular from $(1, 2)$ to $3x+4y = 7$ is $7/(3^2+4^2)^{\frac{1}{2}} = 7/5$.

If the chord length is $2c$, then

$$c^2 = r^2-p^2 = 4-\frac{49}{25} = \frac{51}{25}$$

Hence length of chord is $\frac{2}{5}\sqrt{51}$.

. REVIEW

The review of the problem when we have failed to solve it has been fully discussed in the last section. When we have succeeded in solving the problem, there is a tendency to immediately dismiss the problem and its solution from our mind and start on the next problem. By doing so we ignore a very instructive part of our hard work, unless the problem is a routine one. By reviewing our solution and reconsidering the method that led to that solution we develop our ability to solve problems. Such a review may lead to a much improved solution, and in any case will help our understanding of the method of solution and its possible application to other problems. The importance of checking the results has already been mentioned, and the application of a general result, once proved, to numerical cases is very evident. The search for an alternative solution may not only lead to a shorter solution but also, when successful, enable us to check the result in the numerical 'to find' type of problems. Lastly we try to think of related problems and consider whether the method of proof we used in our problem can be applied to these related problems. We may even try to make up problems to which the procedure used will apply. In other words:

A. CAN WE CHECK THE RESULT?

B. CAN WE FIND A NEATER SOLUTION?

C. CAN WE FIND PROBLEMS WHERE THE RESULT OR THE METHOD OF PROOF CAN BE USED?

These points are illustrated particularly in examples on 'problems and their extensions' given in Chapter 5. Here is one easy illustrative example.

EXAMPLE 11

Given a wooden cube and six pots of paint of different colours, find how many distinct cubes could have every face a different colour.

Solution: If red is one colour, paint the top face red. Then the bottom face can have any one of the five remaining colours. This leaves four colours to be used on the four side faces $= 3!$ ways (since the walls form a closed circuit).

The number of distinct cubes $= 5 \times 3! = 30$.

Restatements:

1. How many distinct dice are possible using the usual numbers 1 to 6.
2. In how many ways may five different vases of flowers be displayed on a rectangular table, one at each corner and one in the centre?

Extensions:

1. Consider the same problem with eight different paints and a regular octahedron, etc.
2. Consider semi-regular solids such as a cube surmounted by a square pyramid.
3. Consider five different paints to paint a cubical box without a lid, painting inside and outside.
4. What is the effect of removing or relaxing the restriction that every face must have a different colour?

In looking over this section on 'what to do when stuck', it may help to collect the various suggestions together.

1. HAVE WE USED ALL THE DATA?

2. CAN WE FIND A SUBSIDIARY PROBLEM?

3. HAVE WE SOLVED A SIMILAR PROBLEM BEFORE?

4. CAN WE RESTATE THE PROBLEM DIFFERENTLY?

5. HAVE WE TRIED 'WORKING BACKWARDS'?

6. HAVE WE CHECKED EACH STEP?

It will be noticed that between the third stage (SUITABLE METHOD) and the fourth stage (SOLUTION) the abilities discussed in Chapter 1 are needed, particularly manipulative

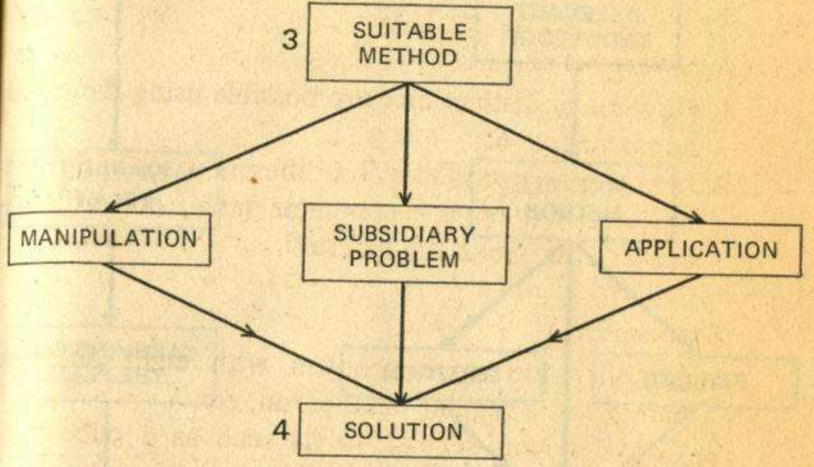


Fig. 2.5

skill. Sometimes the manipulation may be quite complicated, sometimes a little thoughtful application will be required to see a clear path from our subsidiary problem to the desired solution, and occasionally a combination of manipulation and application will be needed to adapt the method chosen to the particular problem under investigation (figure 2.5).

Before ending this chapter, it is of interest to compare our strategy for problem-solving with that given by Professor Polya in his book *How to Solve It* (figure 2.6).

Our strategy arose from the advice given over thirty

years ago to upper-school pupils to help them to solve traditional geometry problems, while Professor Polya's strategy was founded on many years of experience and consideration of general problem-solving. Despite their

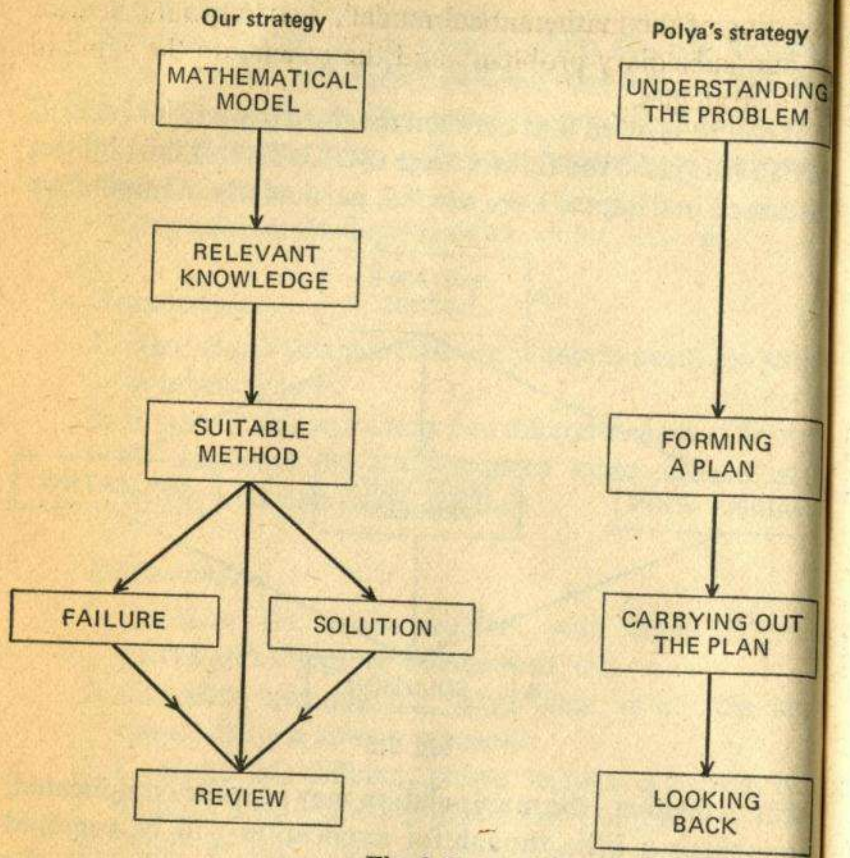


Fig. 2.6

independent development, the final schemes are almost identical.

It is of interest to compare these strategies with the military assessment of strategies:

1. THE AIMS

2. THE PRESENT SITUATION

3. THE IMMEDIATE OBJECTIVE

4. THE METHOD

5. CARRYING OUT THE OPERATION

6. REVIEW OF THE RESULTS.

Here 1 and 2 correspond to our 'formulation and understanding of the mathematical model', 3 corresponds nearest to our 'subsidiary problem', and the others are the same in both strategies.

3. The Inductive Process

There is no single method of problem-solving. The principle of mathematical induction may suffice to establish a certain type of result, once that particular result is known, but it gives no indication of the method by which this particular result was arrived at. There is, however, one general procedure which should help us to discover a possible result of a completely new problem or of an old type of problem in a new and unexpected situation. This procedure is called the 'inductive process', and the 'possible result' will in future be referred to as the 'conjecture'. Not all types of problems lead to conjectures. For example, numerical problems of the 'to find' type usually depend on a standard technique or routine investigation. Real problems, however, as distinct from problems set to illustrate or clarify some particular point or theorem proved in the development of a subject, do usually need some inductive investigation.

The dictionary defines induction as 'the inferring of a general law from particular instances', and the type of reasoning used to make such an inference is called 'inductive reasoning'. Such reasoning is often a particular case of 'plausible reasoning' and many writers refer to it as 'heuristic argument'. It must be emphasized that any conjecture supported only by inductive reasoning may be very controversial. Such a conjecture is regarded as provisional until we have obtained adequate confirmation of its probable truth. This inductive process is described schematically by the flow-chart in figure 3.1, and the rest of the chapter is an expansion of the concepts expressed by this chart together

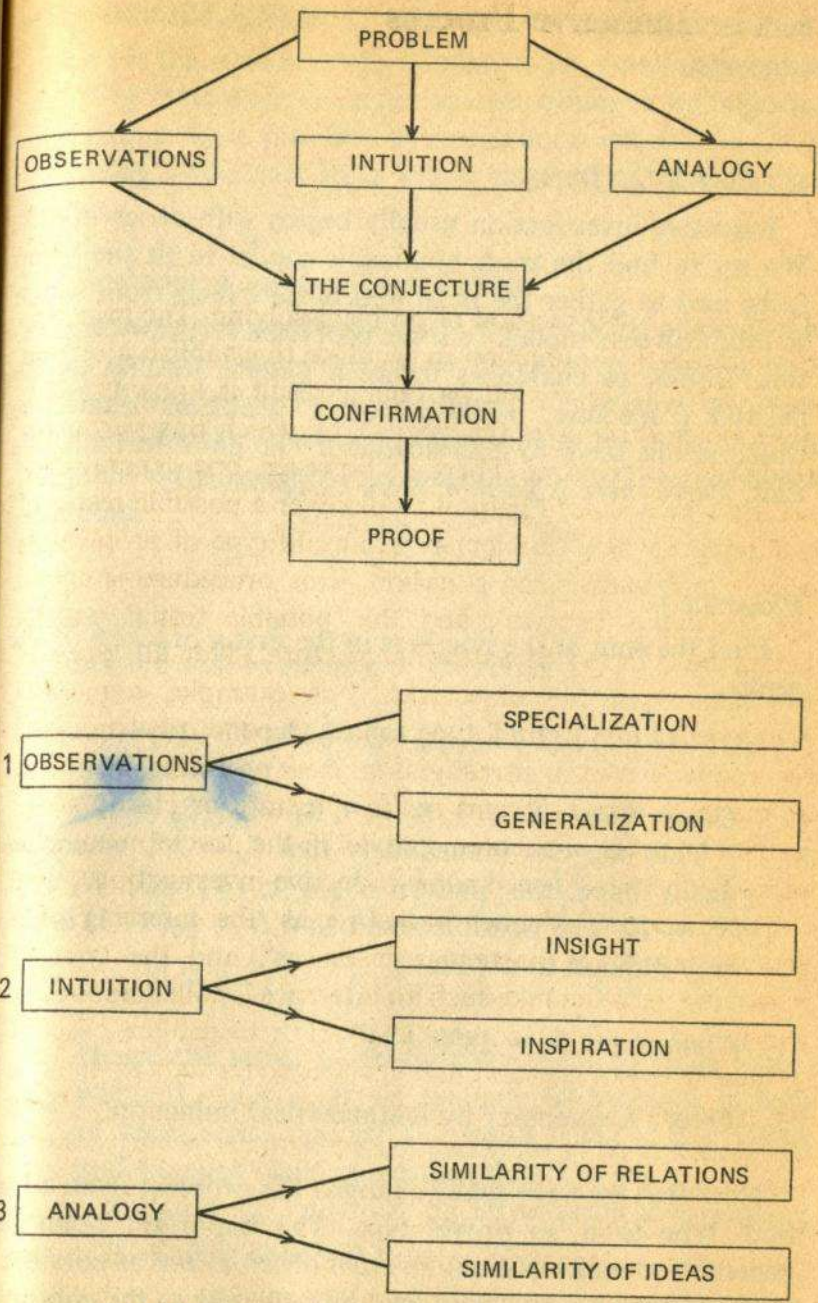


Fig. 3.1. *Representation of the Inductive Process*

with a collection of examples chosen to illustrate these concepts.

1. OBSERVATIONS

Inductive investigation usually begins with observations. We try to find the most applicable model to fit the given facts, and to gather the most appropriate facts from which to establish that model. In some problems very few observations suffice to enable us to see a pattern leading to the desired conjecture, while in other problems numerous observations seem to lead nowhere. The point to realize is that, *where there is pattern, there is significance.*

EXAMPLE 1

Find the sum of the numbers in the first n brackets of the series

$$(1)+(3+5)+(7+9+11)+(13+15+17+19)+\ldots.$$

Observations: $S_1 = 1$, $S_2 = 9$, $S_3 = 36$, $S_4 = 100$.

This suggests immediately that $S_n = N^2$, where we again have the pattern $N_1 = 1$, $N_2 = 3$, $N_3 = 6$, $N_4 = 10$. The possible result that $N = \frac{1}{2}n(n+1)$ is not very difficult to obtain.

Conjecture: $S_n = \frac{1}{4}n^2(n+1)^2$.

Proof: Elementary by mathematical induction.

Note that we have really changed the problem from a 'to find' type to a 'to prove' type. The important point in general is to make as many *relevant* observations as possible. Such relevant observations lead immediately to the concept of *generalization* and *specialization.*

GENERALIZATION

This is the consideration of a larger set containing a given set. The most obvious larger sets are obtained by replacing a constant by a variable, by removing a restriction, or by replacing one object by a whole class which includes that object.

(*a*) Replacing a constant by a variable

Although it seems paradoxical, the replacing of a constant by a variable sometimes leads to an easy solution by a method which was not applicable to the original problem. We now find the answer to the problem by replacing that variable in the result by the constant which the variable had replaced.

EXAMPLE 2

Find the sum to infinity of the series

$$1-\frac{1}{3}+\frac{1}{5}-\frac{1}{7}+\frac{1}{9}+\ldots$$

Observation:

$$S = \left(1-\frac{1}{3}\right)+\left(\frac{1}{5}-\frac{1}{7}\right)+\left(\frac{1}{9}-\frac{1}{11}\right)+\ldots > \frac{2}{3};$$

$$S = 1-\left(\frac{1}{3}-\frac{1}{5}\right)-\left(\frac{1}{7}-\frac{1}{9}\right)-\ldots < 1.$$

Hence the series is clearly convergent to some limit l where $\frac{2}{3} < l < 1$. If we evaluate 60 brackets in each of these series, we get the result $0{\cdot}782 < S < 0{\cdot}788$, and we *might* then conjecture that $S = \frac{1}{4}\pi$.

Suppose we generalize by introducing a variable x. Our original problem is S(1) where

$$S(x) = x-\frac{1}{3}x^3+\frac{1}{5}x^5-\frac{1}{7}x^7+\ldots$$

$$\mathrm{DS}(x) = 1-x^2+x^4-x^6+\dots = \frac{1}{1+x^2} \text{ if } |x| < 1.$$

Hence $\mathrm{S}(x) = \int_0^x \frac{1}{1+t^2}\,dt = [\tan^{-1}t]_0^x = \tan^{-1}x$

(since $\mathrm{S}(0) = 0$).

Hence $\mathrm{S}(1) = \tan^{-1}1 = \frac{1}{4}\pi$.

This proof that $\mathrm{S} = \frac{1}{4}\pi$ would not be valid as it is given because it contains two assumptions which need justification. These are:

(i) Term by term integration of an infinite series.

(ii) What is true *up to* the limit is true *at* the limit.

These assumptions can be rigorously justified, and so the above proof could be made a valid proof.

Example 3

Evaluate $\int_0^\pi \cancel{x-1}\ dx/(2-\cos x)^3$.

Investigation

If we change the integrand into a rational algebraic function by the standard substitution $t = \tan\frac{1}{2}x$, we get

$$2\int_0^\infty \frac{(1+t^2)^2}{(1+3t^2)^3}dt.$$

This is clearly going to be very lengthy and tedious to evaluate by the usual rearrangement into partial fractions.

Suppose, however, we generalize by replacing the constant 2 by a variable λ. This leads us to consider the integral

$$\int_0^\pi \frac{dx}{(\lambda-\cos x)^3}.$$

Relevant knowledge: Such integrals may be evaluated by differentiation under the integral sign with respect to a parameter λ.

Solution:

Consider $\mathrm{F}(\lambda) = \int_0^\pi \frac{dx}{\lambda-\cos x} = \pi/(\lambda^2-1)^{1/2}$ for $\lambda > 1$.

This integral is easy to evaluate by the substitution $t = \tan\frac{1}{2}x$.

$$\mathrm{F}'(\lambda) = -\int_0^\pi \frac{dx}{(\lambda-\cos x)^2} = -\pi\lambda/(\lambda^2-1)^{3/2};$$

$$\mathrm{F}''(\lambda) = 2\int_0^\pi \frac{dx}{(\lambda-\cos x)^3} = \cancel{3\pi\lambda^2/(\lambda^2-1)^{5/2}}.$$
$$\pi(2\lambda^2+1)/(\lambda^2-1)^{5/2}$$

Hence $$\int_0^\pi \frac{dx}{(\lambda-\cos x)^3} = \cancel{3\pi\lambda^2/2(\lambda^2-1)^{5/2}},$$
$$\pi(2\lambda^2+1)/2(\lambda^2-1)^{5/2}$$

and $$\int_0^\pi \frac{\cancel{d\lambda}\ dx}{(2-\cos x)^3} = \cancel{2\pi/3\sqrt{3}}, \text{ when we put } \lambda = 2.$$
$$\pi/2\sqrt{3}$$

(*b*) Removing a restriction

Sometimes a problem is easily solvable when we remove some restriction given in the problem. By studying the solution to the amended problem we may see how the solution for the original problem can be selected from the set of solutions of the more general problem. Here for example is a well-known type of problem encountered in the traditional O-level geometry course.

Example 4

Inscribe a square in a triangle *ABC* so that two of its vertices lie on *BC* and one each on *AB* and *AC*.

Investigation

Suppose we remove the restriction that one vertex lies on AC. It is now quite easy to draw any number of solutions as shown in the left-hand figure. It is *probable* that you will notice that all the fourth vertices lie on a straight line through B. This now suggests a

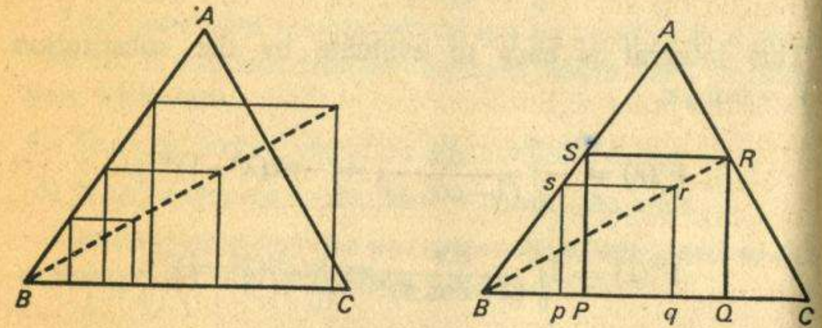


Fig. 3.2

suitable method once we obtain this particular straight line, because the fourth corner which we want will be the intersection of this straight line with the side AC (figure 3.2).

Solution

Draw any square $pqrs$ with p, q on BC and s on BA. Join Br and let it cut AC at R. Draw $RQ \perp BC$ and $RS \parallel BC$, and complete the rectangle $PQRS$ as shown.

To prove: $PQRS$ is a square.

Proof: $\dfrac{rs}{RS} = \dfrac{rB}{RB} = \dfrac{rq}{RQ}$ from the evident similar triangles

$\Rightarrow RS = RQ$, because $rs = rq$ by construction

$\Rightarrow$ *PQRS is a square* (a rectangle with adjacent sides equal).

(c) Replacing one object by a whole class which includes that object

Example 5

The most elegant example is surely Euclid's proof of Pythagoras' Theorem—namely, that in a right-angled triangle the sum of the squares on the two smaller sides equals the square on the largest side.

Inductive process: Euclid looked at the relation $a^2+b^2 = c^2$ and put it in a more general form $ka^2+kb^2 = kc^2$.

He already knew that the areas of similar figures were proportional to the squares on corresponding sides, and he linked this knowledge with the above relations. Squares are regular four-sided polygons, and so the result must be true if he could prove it for any similar polygons with corresponding sides a, b, c. This he saw was simple to prove for similar triangles (figure 3.3).

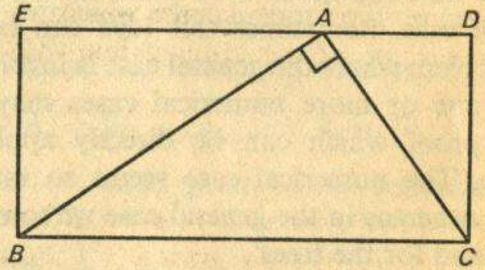


Fig. 3.3

Proof: Triangle ABC has a right angle at A and the rectangle $BCDE$ is completed

$\Rightarrow$ the triangles ABE, ACD, ABC are clearly equiangular and hence similar, with AB, AC, and BC as corresponding sides respectively.

$\triangle ABE + \triangle ACD = \frac{1}{2}\text{rect } BCDE = \triangle ABC$

$\Rightarrow k(AB)^2+k(AC)^2 = k(BC^2)$

$\Rightarrow AB^2+AC^2 = BC^2$, as required.

SPECIALIZATION

This is the converse of generalization, the consideration of a smaller set included in the given set, i.e. the consideration of a subset of the given set and possibly one which contains only one member of the given set. The *replacing of a variable by a constant*, the *introducing of a restriction*, the *replacing of a whole class by a sub-class* are examples of specialization. We also have the cases of extreme specialization, i.e. proceeding to the limiting case when something has a minimum value (often zero) or maximum value (sometimes infinity).

(*a*) Replacing a variable by a constant

This idea, called numerical specialization, is often used in sixth-form teaching and occasionally in university lecturing, especially in first-year courses. For example, general theorems on $n \times n$ determinants (or matrices) are proved by first considering the proof of the simpler 3×3 case, and then pointing out that the method of proof is identical in the general $n \times n$ case. Sometimes this idea can be useful in solving a problem where the general case is involved, in that a study of one or more numerical cases may suggest a method of proof which can be directly applied to the general case. The numerical case seems so much clearer and simpler, whereas in the general case we sometimes 'fail to see the wood for the trees'.

EXAMPLE 6

Show how the $n \times n$ determinant

$$D = \begin{vmatrix} a_{11}+b_{11} & a_{12}+b_{12} & \cdots & a_{1n}+b_{1n} \\ a_{21}+b_{21} & a_{22}+b_{22} & \cdots & a_{2n}+b_{2n} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ a_{n1}+b_{n1} & a_{n2}+b_{n2} & & a_{nn}+b_{nn} \end{vmatrix}$$

can be expanded as the sum of 2^n simpler determinants.

The general case looks very awkward and so we start by considering a special case, $n = 3$. The simplest determinant of this type is

$$D_1 = \begin{vmatrix} a_{11}+b_{11} & a_{12} & a_{13} \\ a_{21}+b_{21} & a_{22} & a_{23} \\ a_{31}+b_{31} & a_{32} & a_{33} \end{vmatrix}.$$

Expanding this down the first column gives:

$$D_1 = (a_{11}+b_{11})\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - (a_{21}+b_{21})\begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix} + $$

$$+ (a_{31}+b_{31})\begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix}$$

$$= \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} + \begin{vmatrix} b_{11} & a_{12} & a_{13} \\ b_{21} & a_{22} & a_{23} \\ b_{31} & a_{32} & a_{33} \end{vmatrix} \quad \text{by definition of a determinant.}$$

This can be written in a much simpler form as

$$|\mathbf{A}_1+\mathbf{B}_1, \mathbf{A}_2, \mathbf{A}_3| = |\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3| + |\mathbf{B}_1, \mathbf{A}_2, \mathbf{A}_3|$$

where vector $\mathbf{A}_r = [a_{1r}, a_{2r}, a_{3r}]$ and $\mathbf{B}_r = [b_{1r}, b_{2r}, b_{3r}]$.

Now consider the most general 3×3 determinant of this type:

$$\begin{aligned} &|\mathbf{A}_1+\mathbf{B}_1, \mathbf{A}_2+\mathbf{B}_2, \mathbf{A}_3+\mathbf{B}_3| \\ &\quad = |\mathbf{A}_1, \mathbf{A}_2+\mathbf{B}_2, \mathbf{A}_3+\mathbf{B}_3| + |\mathbf{B}_1, \mathbf{A}_2+\mathbf{B}_2, \mathbf{A}_3+\mathbf{B}_3| \\ &\quad = |\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3| + |\mathbf{A}_1, \mathbf{B}_2, \mathbf{A}_3| + |\mathbf{A}_1, \mathbf{A}_2, \mathbf{B}_3| + |\mathbf{A}_1, \mathbf{B}_2, \mathbf{B}_3| + \\ &\qquad + |\mathbf{B}_1, \mathbf{A}_2, \mathbf{A}_3| + |\mathbf{B}_1, \mathbf{B}_2, \mathbf{A}_3| + |\mathbf{B}_1, \mathbf{A}_2, \mathbf{B}_3| + |\mathbf{B}_1, \mathbf{B}_2, \mathbf{B}_3| \\ &\quad = \text{sum of } 2^3 \text{ determinants.} \end{aligned}$$

It is evident that exactly the same method of proof will apply to the general $n \times n$ determinant. The actual proof would be tedious but we can see that the expansion of

$$|\mathbf{A}_1+\mathbf{B}_1, \mathbf{A}_2+\mathbf{B}_2, \ldots, \mathbf{A}_n+\mathbf{B}_n|$$

will lead to 2^n determinants in each of which the rth column will be either $\mathbf{A}_r$ or $\mathbf{B}_r$ for $r = 1, 2, \ldots, n$.

(*b*) Introducting a restriction

Teachers of traditional school geometry invariably introduce the 'angle at the centre' theorem by first proving the special case where one arm of the angle passes through the centre, thus introducing a restriction. Students then realize that the method of proof extends directly to the case of no restriction. If, however, the method of proof of the special case did not extend to the general case of no restriction, then this proof would simply constitute a confirmation of the conjecture to be proved.

EXAMPLE 7

Show that the infinite series

$$1+\frac{1}{3}-\frac{1}{2}+\frac{1}{5}+\frac{1}{7}-\frac{1}{4}+\frac{1}{9}+\frac{1}{11}-\frac{1}{6}+\cdots$$

converges and find its sum to infinity.

Investigation

The most direct method is to show that the sum to n terms $S_n \to l$ as $n \to \infty$, because this guarantees the convergence and gives the required sum to infinity simultaneously. We surely notice that the series is made up of triplets of similar terms and this suggests that we put a restriction on n. Instead of $n \in \mathbf{Z}^+$, we will restrict n to be exact multiples of $3 \Rightarrow n = 3m$ where $m \in \mathbf{Z}^+$. Thus we consider the restricted series:

$$S_{3m} = \left(1+\frac{1}{3}-\frac{1}{2}\right)+\left(\frac{1}{5}+\frac{1}{7}-\frac{1}{4}\right)+\cdots+ \\ +\left(\frac{1}{4m-3}+\frac{1}{4m-1}-\frac{1}{2m}\right)$$

$$= \left(1+\frac{1}{3}+\frac{1}{5}+\cdots+\frac{1}{4m-1}\right)- \\ -\left(\frac{1}{2}+\frac{1}{4}+\frac{1}{6}+\cdots+\frac{1}{2m}\right)$$

$$= \left(1+\frac{1}{2}+\frac{1}{3}+\frac{1}{4}+\cdots+\frac{1}{4m}\right)- \\ -\left(\frac{1}{2}+\frac{1}{4}+\cdots+\frac{1}{4m}\right)-\left(\frac{1}{2}+\frac{1}{4}+\cdots+\frac{1}{2m}\right)$$

$$= \sum_{1}^{4m}\frac{1}{r}-\frac{1}{2}\sum_{1}^{2m}\frac{1}{r}-\frac{1}{2}\sum_{1}^{m}\frac{1}{r}.$$

We may know the important result that

$$\sum_{1}^{n}\frac{1}{r} \to \log n+\gamma,$$

where γ is Euler's constant. Hence, as $m \to \infty$,

$$S_{3m} \to (\log 4m+\gamma)-\tfrac{1}{2}(\log 2m+\gamma)-\tfrac{1}{2}(\log m+\gamma)$$

$$= \frac{3}{2}\log 2.$$

It is easy to see that $S_{3m\pm 1} \to S_{3m}$ as $m \to \infty$ and so we have

$$S_n \to \frac{3}{2}\log 2 \text{ for all } n \in \mathbf{Z}^+.$$

(*c*) Replacing a whole class by a sub-class

The classical proofs of some of the theorems in analysis, particularly in the theory of analytical function, involve the replacing of a whole class of simply connected curves by just one member of that class. The more common use, however, of this type of specialization is in making a confirmation of a conjecture already suggested. For example, Apollonius' Theorem does not seem a reasonable proposition to a class when they are first introduced to it. A good teacher would probably ask his pupils to prove this result first for some special cases.

EXAMPLE 8

If D is the mid-point of sides BC of any triangle ABC, prove that

$$AB^2+AC^2 = 2AD^2+2BD^2.$$

Investigation: Suppose we try some special cases.

Sub-class of isosceles triangles.

In figure 3.4(a) the median bisects the base at right angles,

hence $AB^2+AC^2 = 2x^2 = 2(y^2+h^2) = 2y^2+2h^2$

$\Rightarrow \quad AB^2+AC^2 = 2AD^2+2BD^2.$

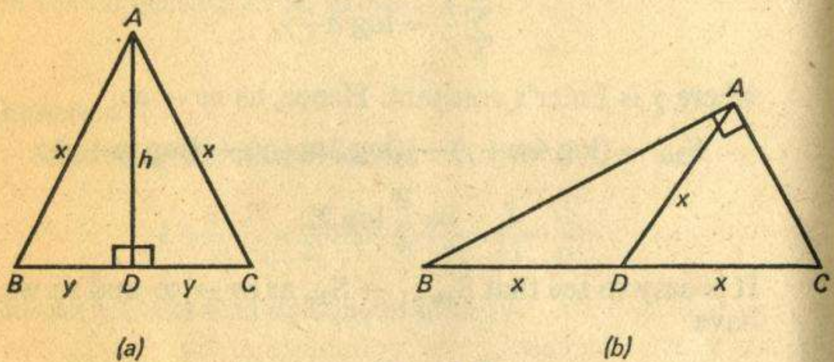


Fig. 3.4

Sub-class of right-angled triangles.

In figure 3.4(b) the median AD has length $\frac{1}{2}BC$,

hence $\quad AB^2+AC^2 = BC^2$ (Pythagoras) $= 4x^2$

$$= 2AD^2+2BD^2.$$

In each case the proof is very easy and, as a result, the pupils will be quite happy to accept the result as a probable proposition for the general triangle.

(d) Extreme specialization

The testing of an extreme case is analogous to the process of proceeding to the limit, and is often helpful both as an

immediate confirmation of a proposed conjecture and sometimes in the forming of the conjecture. For example, is there any connection between the length of the line joining the mid-points of the two non-parallel sides of a trapezium and the lengths of the parallel sides?

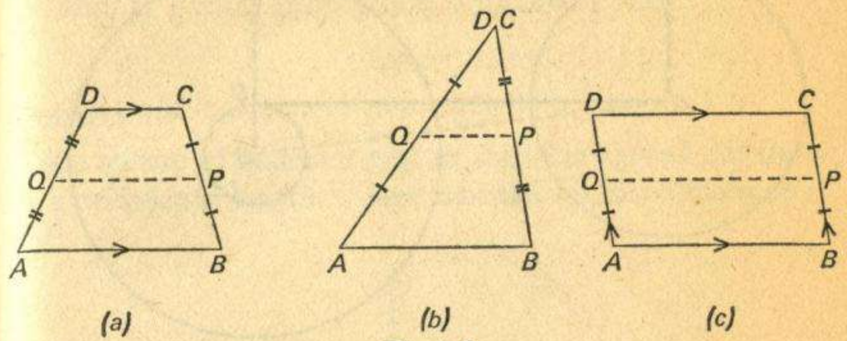


Fig. 3.5

We look at the two extreme cases of a trapezium—when it becomes a triangle or a parallelogram. In the triangle in figure 3.5(b) we know that $PQ = \frac{1}{2}AB$ by the 'mid-points' theorem; in the parallelogram in figure 3.5(c) we see that $PQ = AB = CD$. It is not very hard to make the conjecture $PQ = \frac{1}{2}(AB+CD)$, nor to notice that, since $PQ \parallel AB$ in both the extreme cases, it is probable that this is also true for the general trapezium.

EXAMPLE 9

Construct a direct common tangent to two unequal circles.

Investigation

We first consider the extreme case when one of the circles becomes a point. This, of course, gives us the standard known construction of a tangent from a point to a circle. We might then consider what happens as we shrink the smaller circle to a point. Sooner or later we would realize that, as shown in figure 3.6,

if we shrink the two circles by equal amounts, the direction of the common tangent is always parallel to its original direction and so does not change. Hence,

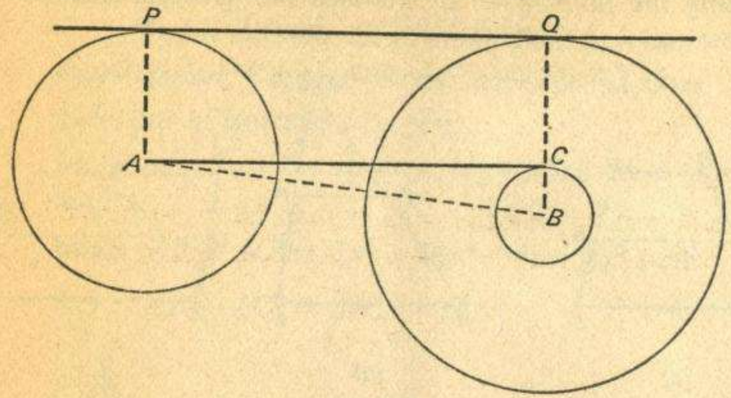


Fig. 3.6

as we shrink down to a point A, the tangent direction is still the one required for our direct common tangent to the two circles.

Solution

Draw circle centre B and radius $r_B - r_A$. Draw as usual the tangent AC from A to this circle to touch at C. Join BC and produce to cut original circle at Q. Draw AP parallel to BQ.

Then PQ is the required direct common tangent.

2. INTUITION

Intuition is the 'power of direct immediate perception' or 'unreasoned knowledge'. Intuition tends to prejudice us in favour of some results and against other quite similar results. For example, the proposition that 'of all surfaces of equal volume the sphere has the minimum area' gets our immediate intuitive support. We instinctively think of raindrops, soap bubbles, and planets, and thus intuitively feel that the sphere is favoured by nature herself. Yet this

intuition can let us down. For example, the proposition that 'if a sphere is pierced coaxially by a cylinder of length l, the volume remaining is $\frac{1}{6}\pi l^3$', does not receive our intuitive support. We instinctively expect that the radius of the sphere must be an essential parameter in the volume remaining. Yet the proposition is a correct one.

EXAMPLE 10

If a sphere of diameter greater than l is pierced coaxially by a cylinder of length l, show that the volume remaining is $\frac{1}{6}\pi l^3$.

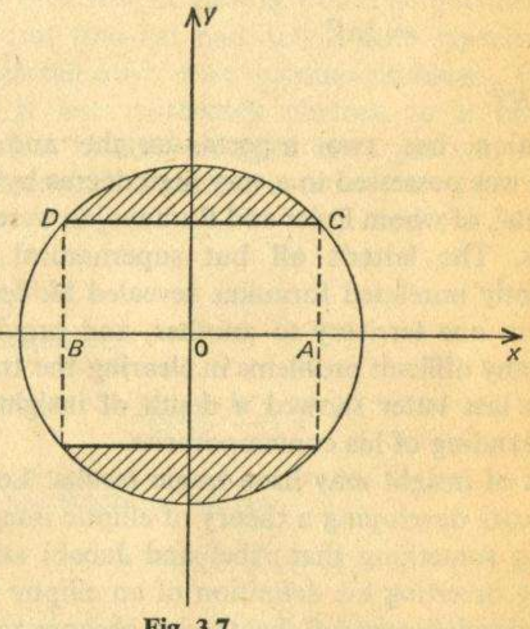


Fig. 3.7

Investigation

Consider the two extreme cases.

$l = 0 \Rightarrow V = 0$ which holds for $V = \frac{1}{6}\pi l^3$.

$l = 2R \Rightarrow V = \frac{4}{3}\pi R^3$ which again holds for $V = \frac{1}{6}\pi l^3$.

Hence the result is confirmed in two cases (figure 3.7).

Solution

Volume V remaining is the difference between the frustum of a sphere of thickness AB and the cylinder of length AB and radius AC. We know the standard method for volumes of revolution and here we have the circle $x^2+y^2 = R^2$,

$$V = \int_{0B}^{0A} \pi y^2 dx - \pi(AC)^2 . AB$$

$$= \pi \int_{-\frac{1}{2}l}^{\frac{1}{2}l} (R^2-x^2)dx - \pi(R^2-\tfrac{1}{4}l^2) . l$$

$$= 2\pi[R^2x-\tfrac{1}{3}x^2]_0^{\frac{1}{2}l} - \pi(R^2l-\tfrac{1}{4}l^3)$$

$$= \tfrac{1}{6}\pi l^3.$$

INSIGHT

Intuition has two aspects—insight and inspiration. Insight was possessed to a very high degree by the so-called 'algorists', of whom Euler and Ramanujan were the supreme masters. The latter's all but supernatural insight into apparently unrelated formulae revealed hidden trails leading from one territory to another, and provided analysts with many difficult problems in clearing the trails. Galois's famous last letter showed a depth of insight beyond the understanding of his contemporaries.

Lack of insight may have tragic results. Legendre spent forty years developing a theory of elliptic integrals without noticing something that Abel and Jacobi saw at once—that, by inverting his definition of an elliptic function, the whole development of the theory became much simpler. Legendre defined an elliptic function as $y = F(x)$ where

$$y = \int_0^x \frac{dt}{\{(1-t^2)(1-k^2t^2)\}^{\frac{1}{2}}};\ k^2 < 1.$$

Jacobi suggested that this integral should be used to define x as a function of y, i.e. $x = f(y)$, written nowadays as

$x = \text{sn}\, y$. We have a similar situation with the circular functions. Instead of using the integral

$$y = \int_0^x \frac{dt}{(1-t^2)^{\frac{1}{2}}}$$

to define y as a function of x, we use it to define x as a function of y, written as $x = \sin y$. We all realize how much easier it is to work with $\sin x$ than with the inverse function $\sin^{-1} x$.

Insight cannot be taught. It can be acquired only by experience. In Chapter 1 we used the term 'application' to mean a very little insight such as would be possessed by an average student who has had only a little experience with solving problems other than routine problems. Once the application is not *reasonably* obvious to a student of mathematics, it becomes insight. Here are two examples where simple insight is needed.

Example 11

Solve the equation

$$\frac{1}{x^2}+\frac{1}{(2-x)^2} = 1.$$

On clearing the fractions we get

$$x^4-4x^3+2x^2+4x-4 = 0.$$

Since $x = \pm 1, \pm 2, \pm 4$, do not fit the equation, there are no integer roots. Hence we have the awkward task of finding two quadratic factors of the quadric. But, a closer study of the original equation plus a little insight gives an easy solution. What can we do to the expressions

$$\frac{1}{x^2} \text{ and } \frac{1}{(2-x)^2}$$

to make them more convenient for manipulative purposes? Surely it will be helpful to change them as follows:

put $\quad x = 1-y \Rightarrow \dfrac{1}{(1-y)^2}+\dfrac{1}{(1+y)^2} = 1$

giving $\quad y^4-4y^2-1 = 0$, which is a quadratic in y^2.

Hence $\quad y^2 = 2\pm\sqrt{5} \Rightarrow y = \pm(2\pm\sqrt{5})^{\frac{1}{2}}$,

and $\quad x = 1\pm(2\pm\sqrt{5})^{\frac{1}{2}}$.

EXAMPLE 12

Evaluate the integral

$$I = \int_0^{\frac{1}{2}\pi} \frac{\sin^3 x}{\sin x+\cos x}\,dx.$$

Investigation

We realize that the usual substitutions will lead to very awkward integrands. Sooner or later we *intuitively* see that the same answer will be given by

$$\int_0^{\frac{1}{2}\pi} \frac{\cos^3 x}{\sin x+\cos x}\,dx.$$

The justification for our intuition is immediately given by the simple substitution $x \to \frac{1}{2}\pi-x$. Now a little insight is needed—each is awkward to evaluate separately, but the sum of the two is very easy to find.

Solution

$$2I = \int_0^{\frac{1}{2}\pi} \frac{\sin^3 x+\cos^3 x}{\sin x+\cos x}\,dx = \int_0^{\frac{1}{2}\pi} (1-\tfrac{1}{2}\sin 2x)dx$$

$$= [x+\tfrac{1}{4}\cos 2x]_0^{\frac{1}{2}\pi} = \frac{\pi}{2}-\tfrac{1}{2}.$$

Hence $I = \frac{1}{4}(\pi-2)$.

Some writers denigrate such examples as 'slick tricks'. It would be interesting to learn how they acquired the insight so necessary to the solution of unusual problems. Depth of insight comes from experience with a large variety of practical and theoretical problems which are not of routine type. Whereas a competent mathematician needs and uses this aspect of intuition, a creative mathematician needs something further—he needs also 'inspiration'.

INSPIRATION

Inspiration is an 'inspired guess'. Poincaré once studied this occurrence of inspiration in creative mathematics and his conclusion was that 'mathematical discoveries are never born of spontaneous generation; they always presuppose a soil seeded with preliminary knowledge and well-prepared by labour'. Edison certainly agreed with him when he said: 'genius is 1 per cent inspiration and 99 per cent perspiration'. Without drudgery and a flash of inspiration few great discoveries have been made. Intuition and the subconscious mind work only on past experience which furnishes the raw material for the flash of inspiration.

With really tough problems we go through two phases—a quiet phase and an active phase. In the active phase we work hard at the problem and try to make some progress, however small. It may be that we can prove a special case of the problem, or find a subsidiary problem that is more accessible, or discover a related problem whose solution gives us confirmation that our problem is solvable. In the quiet phase we do little practical work connected with the problem, but we recall it from time to time until quite suddenly an apparently random thought suggests a new way of looking at the problem, or an analogue of the problem, or an extension of the problem that seems to be more accessible. This now returns us to the active phase and so on. Hamilton laboured fifteen years to invent an algebra

of rotation in three dimensions until, while walking one day in the country, he suddenly got an inspired thought that $a \,.\, b \neq b \,.\, a$ in the algebra he was seeking. So started his famous *theory of quaternions*. Archimedes in his bath, Hamilton on his walk, and Einstein in his study, all had these sudden 'inspired guesses'.

At the level for which this book is written we do not expect to meet problems needing such inspired guesses for progress with a solution. It is of interest, however, to consider the following problem in the light of what is written above.

EXAMPLE 13

A sequence $\{U_n\}$ is defined by the recurrence relation

$$U_{n+1} = U_n - U_n^2;\ U_1 = \tfrac{1}{2}.$$

Prove that $U_n \to \dfrac{1}{n}$ as $n \to \infty$.

Investigation

We easily establish properties that we would expect to lead to the required result, but in every case the same type of inequality comes up. Properties such as:

(i) Sequence $\{U_n\}$ is monotone decreasing.

(ii) $U_n < \dfrac{1}{n}$ and so $U_n \to 0$ as $n \to \infty$.

(iii) Sequence $\{nU_n\}$ is monotone increasing.

(iv) $nU_n \to l \leqslant 1$ as $n \to \infty$.

We always seem to finish up with inequality $nU_n < 1$, whereas we want to get nU_n greater than something.

i.e., $f(n) < nU_n < 1$.

The function $f(n) = 1 - n^{-\frac{1}{2}}$ works, but is there any

reason why we should think of such a function? Here is one proof of first year university standard.

(i) $U_{n+1} - U_n = -U_n^2 \Rightarrow \{U_n\}$ is monotone decreasing.

(ii) $U_n < \dfrac{1}{n+1}$ by induction, since true for $n = 2$ and

$$U_{k+1} = U_k(1-U_k) < \frac{1}{k+1}(1-U_{k+1}) \text{ since } U_{k+1} < U_k$$

$$\Rightarrow U_{k+1}\left(1+\frac{1}{k+1}\right) < \frac{1}{k+1} \Rightarrow U_{k+1} < \frac{1}{k+2}.$$

(iii) hence $\dfrac{1}{nU_n} > 1+\dfrac{1}{n}$ and $\dfrac{1}{1-U_n} < 1+\dfrac{1}{n}$ $(n > 1)$.

(iv) Also $\dfrac{1}{U_{k+1}} = \dfrac{1}{U_k} + \dfrac{1}{1-U_k} \Rightarrow \dfrac{1}{U_{k+1}} - \dfrac{1}{U_k} < 1+\dfrac{1}{k}$.

Write this down for $k = (n-1), (n-2), \ldots, 1$ and add:

$$\frac{1}{U_n} - \frac{1}{U_1} < (n-1) + \sum_1^{n-1} \frac{1}{r} \Rightarrow \frac{1}{U_n} < (n+1) + \sum_1^{n-1} \frac{1}{r},$$

$$\text{hence } 1+\frac{1}{n} < \frac{1}{nU_n} < 1+\left(1+\sum_1^{n-1}\frac{1}{r}\right)\Big/n.$$

L.H.S. and R.H.S. $\to 1$ as $n \to \infty$,

$$\text{hence } \frac{1}{nU_n} \to 1 \Rightarrow U_n \to \frac{1}{n}.$$

3. ANALOGY

We think of analogy as a 'similarity in structure' between two systems. We perceive a one–one correspondence between the system in an original problem and a new system, often in a completely different context. We then try to find out

how good this correspondence is and how far the correspondence extends. We think of the different degrees of similarity as:

(*a*) Identical structures—*isomorphisms.*

(*b*) Similar structures—*homomorphisms.*

(*c*) Vaguely similar structures.

Such analogues can be considered in two ways—similarity of relations and similarity of ideas.

SIMILARITY OF RELATIONS

Here our two systems are analogous because they agree in clearly definable relations of their respective parts. When the analogy is vague, we must try to clarify it, to improve it and to extend it. When the analogy is clear, the correspondence has probably been well explored already. Some of the well-known analogues frequently used are:

(i) *Plane and solid geometry.*

(ii) *Numbers and figures.*

(iii) *The finite and the infinite.*

(iv) *Infinite series and integrals.*

(v) *Projectiles and planetary motion.*

Analogues of such types often play some part in any important discovery.

EXAMPLE 14

Bernoulli's problem—evaluate $\sum_{1}^{\infty}\frac{1}{r^2}$.

Investigation

Euler looked at two known results:

$$\sin x = x-\frac{1}{3!}x^3+\frac{1}{5!}x^5-\frac{1}{7!}x^7+\ldots, \qquad (1)$$

$$\sin x = 0 \text{ has roots } 0, \pm\pi, \pm 2\pi, \pm 3\pi, \ldots. \qquad (2)$$

The second one caused him to think of the identity

$$f(x) \equiv k(x-a_1)(x-a_2)\ldots(x-a_n), \qquad (3)$$

where the a_r are the roots of the *polynomial* equation $f(x) = 0$.

Now he saw two analogues, one between the algebraic equation $f(x) = 0$ and the transcendental equation $\sin x = 0$, the other between the finite and the infinite. He decided to apply result (3) with an infinite number of roots to the equation (2). Thus,

$$\sin x = x\left(1-\frac{x^2}{\pi^2}\right)\left(1-\frac{x^2}{4\pi^2}\right)\left(1-\frac{x^2}{9\pi^2}\right)\ldots \qquad (4)$$

Now by equating the coefficients of x in the *known* expansion (1) and the *unproved* conjecture (4) he got

$$-\frac{1}{6} = -\frac{1}{\pi^2}-\frac{1}{4\pi^2}-\frac{1}{9\pi^2}-\ldots,$$

$$\text{hence } \frac{\pi^2}{6} = 1+\frac{1}{4}+\frac{1}{9}+\ldots = \sum_{1}^{\infty}\frac{1}{r^2}.$$

Euler had applied a rule for algebraic equations to transcendental equations and a rule for the finite to the infinite. These were very daring analogues in 1750 and he realized the pitfalls in his method. Not until he had established very strong confirmation of the probable truth of this conjecture did he publish the result that

$$\sum_{1}^{\infty}\frac{1}{r^2} = \frac{1}{6}\pi^2.$$

EXAMPLE 15

Find any restrictions on the constants a, b, c such that the equations

$$ax+by+c = 0, \qquad x^2+y^2 = 1$$

possess real roots.

Investigation

This problem is quite easy to solve in the algebraic context, but it is of interest to consider its analogue in the geometric context. The first equation represents a straight line and the second equation a circle of centre (0, 0) and radius 1. Real roots will correspond to real points of intersection. The algebraic and geometric systems of this problem are isomorphic systems and so the proof from one system simply needs to be given its appropriate interpretation in the other system.

Solution

The circle and straight line clearly intersect in real points if the perpendicular distance from the centre to the line is $\leqslant$ the radius.

Hence $\left|\dfrac{c}{\sqrt{(a^2+b^2)}}\right| \leqslant 1 \Rightarrow c^2 \leqslant a^2+b^2.$

SIMILARITY OF IDEAS

Here a problem is seen in two different contexts such as:

(i) *Algebraic and geometric.*

(ii) *Optical and shortest distance problems.*

We perceive two different interpretations of the one problem and sometimes find the problem is much easier to solve in one context than in the other. Sometimes known properties in the new context suggest a way of tackling the problem in its original context. It should be realized that the proofs of the analogous properties in the two contexts will not be identical, partly because of the differences in the physical implications of these properties in their respective contexts. Here is a well-known geometrical problem, called 'Schwartz's problem', which is very difficult to solve without first considering the analogous problem in an optical context. The optical property that links shortest distance properties

in geometry with rays of light in optics is Fermat's 'principle of least time'—that a ray of light passing from point A to point B always traverses the path of least time.

Example 16

The main land-drains for a level piece of ground are in the shape of an acute angle triangle ABC. Three junctions are to be constructed, one on each side, so that each pair can be joined by a subsidiary main for emergencies. If the contract cost is a fixed amount per metre, where should the junctions be situated so that the total cost of the three subsidiary drains is minimized?

Mathematical model

A triangle ABC with points P, Q, R chosen, one on each side.

Comprehension

We seek the positions of P, Q, R so that the perimeter of triangle PQR is a minimum.

Investigation

Suppose we consider the analogue of our problem in an optical context. We consider the sides to be mirrors. If K is a point on QR, then a ray of light from K will return to K after reflection in each of the mirrors in turn along the path of least time. Hence $KQPRK$ will be a path of least length since we have a uniform medium in which the light travels. Now we know a property in the optical context that helps us to find the desired triangle PQR—the law of reflection. This says that on reflection the incident and reflected rays are equally inclined to the normal at the point of reflection. Hence each pair of rays (such as QP and PR) are equally inclined to the appropriate mirror.

Thus we seek a triangle PQR such that its sides are equally inclined to the sides of triangle ABC at points P, Q, and R respectively. We now return to our original geometrical context.

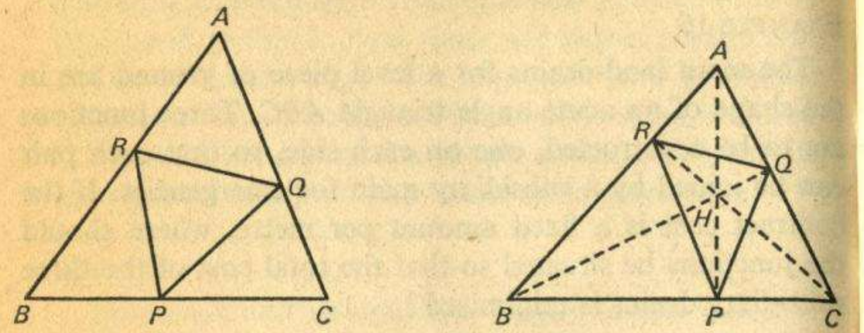


Fig. 3.8

Solution

We know that the pedal triangle (the triangle whose vertices are the feet of the altitudes of triangle ABC) has the property that each pair of sides are equally inclined to the side they meet. Students should establish this property by using the cyclic quadrilaterals $BPHR$ and $CPHQ$. We therefore make the following assertion: *the points P, Q, R are the feet of the perpendiculars from A, B, C to the opposite sides.*

Now to complete the proof in the geometrical context we must prove that the perimeter of the pedal triangle is less than the perimeter of any other inscribed triangle. The easy proof of this is left as an exercise for the reader. Notice that this proof would not suffice in the optical context, where to complete the proof we must show that the pedal triangle is the only inscribed triangle with the 'equally inclined to sides' property required for the law of reflection to be satisfied.

ASSUMPTIONS

When working inductively on mathematical problems we often need to make assumptions which later have to be justified. The main assumption we make is that our problem is not an exceptional one, i.e., that a solution exists, that our functions are 'well-behaved', and that 'proceeding formally' with standard routine operations can be carried out to help us form a conjecture for our problem. The process by which we often arrive at our conjecture is really a 'heuristic' type of process, and so our conjecture will be only provisional until we supply a proof in which any heuristic assumptions used are justified by rigorous analysis.

The commonest assumptions are:

(i) A system of n linear equations with n unknown is solvable.

(ii) Formal operations such as the differentiation and integration of infinite series can be carried out.

(iii) Limit operations are commutative, such as

$$\sum_{n=1}^{\infty} \int_a^b f_n(x)dx = \int_a^b \sum_1^{\infty} f_n(x)dx.$$

(iv) In general our functions are well-behaved and possess a power series expansion with the property that the greater the number of terms of this series are used, the better the approximation to the function.

There are many other less common assumptions and one of the principal assets of a mathematician is to know when to make such assumptions and how far to trust them. This only comes with experience in general and self-involvement in particular. For example, if we take the length of a continuous plane curve as the limit of the length of the inscribed polygonal lines, we do get a correct result, but if

we take the area of a three-dimensional surface as the limit of the area of inscribed polyhedra, we sometimes find that our results vary according to the type of inscribed polyhedra used.

Example 17

Here are some simple cases to show that 'proceeding formally' with standard mathematical operations may lead to disaster.

(*a*) Find the sum to infinity of the series

$$1+2+4+8+16+\ldots$$

Solution: $S = 1+2+4+8+16+\ldots$

$$2S = 2+4+8+16+\ldots = S-1,$$

hence $S = -1$ (which shows the sum of positive terms is negative).

(*b*) Evaluate $\int_0^\pi \theta \cos\theta \, d\theta$

Solution

Put $\sin\theta = t \Rightarrow \cos\theta d\theta = dt$ and $\theta = 0$ or π gives $t = 0$.

$$\int_0^\pi \theta\cos\theta \, d\theta = \int_0^0 \sin^{-1} t \, dt = 0.$$

(*c*) Solve the equation $\tan 3\theta + \cot\theta = 0$

Solution

$$\cot\theta + \tan(2\theta+\theta) = 0$$

$$\cot\theta + \frac{\tan 2\theta + \tan\theta}{1-\tan 2\theta\tan\theta} = 0$$

$$\cot\theta - \tan 2\theta - \tan 2\theta + \tan\theta = 0$$

$$\cot\theta + \tan\theta = 0 = (1+\tan^2\theta)/\tan\theta$$

$$\cot\theta + \tan\theta = 0 = (1+\tan^2\theta)/\tan\theta$$

$$\text{hence } \tan^2\theta = -1, \text{ since } \tan\theta \neq 0,$$

$$\tan\theta = \pm i.$$

Note: Since the first series is divergent and the second integral would give the same answer if θ were replaced by $f(\theta)$, the errors in (*a*) and (*b*) are easy to discover. The last part is not so elementary although we realize an error must have been made since $\theta = \frac{1}{4}\pi$ is clearly a solution of the equation.

NOTATION

After Newton discovered the calculus its rapid development was helped enormously by the notation devised by Leibniz. It is sometimes difficult to explain a problem's logical structure without devising a suitable notation or some other form of natural aid. Sometimes this natural aid is the formation of an unusual mathematical model such as the buttons and string model for the 'four knights' problem in Chapter 2, or the use of matrices with vacant cells to enable us to group conveniently the facts in the 'Smith–Jones–Robinson' type of brain-teaser. Sometimes when the problem is completely new to us, we have to devise a simple suitable notation that will group together the facts of the problem and significant ideas about them. In an ideal notation the order and connection of the notation should suggest the order and connection between the objects or elements of our problem. Above all, a good notation should be as simple as possible, easy to remember and unambiguous. Such notation may be very complicated

in more advanced mathematics, such as *tensors* where both the affixes and the suffixes have order and meaning. Two examples often found in textbooks are:

(*a*) The notation N_r for the number of elements of dimension r possessed by a polytope: this is simple and unambiguous with N_0, N_1, N_2, . . ., representing the number of vertices, edges, faces, . . ., of the polytope.

(*b*) The notation α for a plane, $\alpha\beta$ for the intersection of planes α and β, $\alpha\beta\gamma$ for the intersection of planes α, β, and γ: this is a very ambiguous notation because $\alpha\beta\gamma$ could represent a point, or a straight line or a set of two or three parallel straight lines.

Example 18

Consider the product of two $n\times n$ matrices **A** and **B**. We find it impossible to work with the full definitee

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

but very easy to carry on with the simple notation

$$\mathbf{A} = [a_{rs}].$$

Here a_{rs} is the element in the rth row and sth column.

Let $\mathbf{AB} = [c_{rs}]$

where c_{rs} is the element in the rth row and sth column of the product of **A** and **B**, i.e.

c_{rs} = product of rth row of **A** with the sth column of **B**

$$= [a_{r1}a_{r2}\ldots a_{rn}]\begin{bmatrix} b_{1s} \\ b_{2s} \\ \cdot \\ \cdot \\ \cdot \\ b_{ns} \end{bmatrix} = \sum_{i=1}^{n} a_{ri}b_{is}.$$

Hence $\mathbf{AB} = [a_{rs}][b_{rs}] = \left[\sum_{i=1}^{n} a_{ri}b_{is}\right].$

Example 19. *The abridged notation*

In coordinate geometry the notation $u = 0$ is used to represent a straight line and $S = 0$ to represent a conic. This enables quite general results to be written down in very simple form. Suppose $u = 0$ and $v = 0$ are distinct straight lines and $S = 0$ is any conic which has $t = 0$ as a tangent.

1. $u+kv = 0$ represents any straight line passing through the intersection of $u = 0$ with $v = 0$. By giving k the appropriate value we can find the line also satisfying some other condition.
2. $S+kuv = 0$ is any conic passing through all the intersections of the two lines with the conic.
3. $S+kt = 0$ is any conic that touches the given conic at the same point that the tangent $t = 0$ touches it.
4. $S+kut = 0$ is any conic that touches the given conic at the same point as the tangent $t = 0$ does, and also passes through the intersections of the line $u = 0$ with the given conic.
5. If $u = 0$, $v = 0$, and $w = 0$ are three sides of a triangle, then the equation $uv+\lambda vw+\mu wu = 0$ represents any conic passing through the vertices of the triangle. By finding the appropriate values of the constants λ and μ,

we can find the equation of the circumcircle of the triangle. This is done by making the coefficients of the term in xy zero and making the coefficients of x^2 and y^2 equal.

CONFIRMATION

When our inductive investigation leads us to a conjecture which seems reasonable, we try to prove it is a correct one. If we fail to find a proof, it will reinforce our belief in the conjecture if we can find some further confirmation of it. The examination of further numerical cases does not really strengthen our belief in the conjecture.

EXAMPLE 20

We conjecture $n^2-79n+1{,}601$ is a ~~pure~~ prime number for $n \in \mathbf{Z}^+$.

Investigation

The conjecture holds for $n = 1, 2, 3, 4$, and many further values of n. Yet the proposition fails when $n = 80$:

$$80^2-79(80)+1{,}601 = 1{,}681 = (41)^2.$$

This single counter-example is sufficient to prove the conjecture is incorrect. A thoughtful student who examined the conjecture closely before embarking on manipulative work would realize immediately that $n = 1{,}601$ gives a counter-example because the expression would then be clearly divisible by 1,601.

However, our belief in the conjecture would be greatly strengthened if we could:

(i) *Confirm some other consequence of the conjecture.*

(ii) *Obtain a known result as a consequence of the conjecture.*

The more this new consequence differs from the original verified conjecture, the greater will be our confidence in that conjecture.

EXAMPLE 21

We conjecture that the area of a cyclic quadrilateral $ABCD$ is

$$A = \{(s-a)(s-b)(s-c)(s-d)\}^{\frac{1}{2}},$$

where $2s = a+b+c+d$.

Investigation

The dimension of A is $4^{\frac{1}{2}} = 2$ which is correct for an area. The conjecture is symmetric in the four sides a, b, c, d which is obviously a necessary requirement of the area of the quadrilateral. We require further confirmation and we try the extreme cases and any accessible special cases.

Extreme cases

If two vertices, say A and D, coincide, then $d = 0$ and the quadrilateral degenerates into a triangle. The conjecture gives

$$A = \{s(s-a)(s-b)(s-c)\}^{\frac{1}{2}}$$

which is the correct result for a triangle.

If three vertices coincide (say B, C, D) then $s = a$ and the quadrilateral degenerates into the straight line AB. The conjecture gives $A = 0$ which is the correct result.

Special cases

If $a = b = c = d$, then $s = 2a$ and the conjecture gives $A = a^2$. The only quadrilaterals with four equal sides are squares and rhombi. The only such quadrilateral that can be inscribed in a circle is a square, for which the correct area is $A = a^2$.

If $a = c$, $b = d$, then $s = a+b$ and the conjecture gives the value $A = ab$. The only quadrilaterals with opposite sides equal are parallelograms and rectangles.

The only such quadrilaterals inscribable in a circle are rectangles for which the correct area is $A = ab$.

These verifications do not establish the conjecture, but they do greatly strengthen our belief in its validity.

PROOF

The last stage in our over-all strategy is *proof*. Methods of proof are discussed in great detail in the next chapter. We should, however, notice that the proof may arise out of the inductive procedure in two ways, by examination of the general consequence and by rigorous justification of the inductive procedure used to establish the conjecture. Examination of the general consequence has occasionally been used by A-level students acquainted with the inductive process to answer an examination question where they have failed to see the method intended by the examiner.

EXAMPLE 22

Find the value of

$$P_n = \left(1-\frac{4}{1}\right)\left(1-\frac{4}{9}\right)\left(1-\frac{4}{25}\right)\dots\left(1-\frac{4}{(2n-1)^2}\right).$$

Investigation

They considered the special cases to guess a conjecture

$$P_1 = -\frac{3}{1}, P_2 = -\frac{5}{3}, P_3 = -\frac{7}{5}.$$

The conjecture $P_n = -\dfrac{2n+1}{2n-1}$ seems obvious.

Confirmation

$P_4 = -\dfrac{7}{5}\times\dfrac{45}{49} = -\dfrac{9}{7}$ which equals P_4.

This conjecture was now proved quite easily by mathematical induction.

Before ending this discussion of the inductive process, I must mention the greatest discovery by the inductive process —Mendeloff's *periodic law of the elements*—whose truth was established many years later by means of the quantum theory. There have been many famous inductive conjectures which even today have not been proved or disproved. The two best-known are *Fermat's Last Theorem* and *Goldbach's Theorem*:

Fermat's Last Theorem is that 'the equation $x^n+y^n = z^n$ has no positive integer solutions for x, y, z if $n > 2$'.

Goldbach's Theorem is that 'every even number greater than 4 is the sum of two odd primes'.

4. Methods of Proof

When we first encountered the idea of a precise proof in our traditional school geometry course, we were usually considering the end-product of a deductive process. We learned then that every step in our proof had to be justified by deductive reasoning, and that the steps themselves had to be self-consistent. Later in proving the converses of some of the propositions already proved by deduction we were shown new methods of proof by *reductio ad absurdum* and 'proof by exhaustion' (which sometimes had the same effect on the pupils!). Then in the sixth form we encountered a new type of proof—proof by 'mathematical induction'.

While in the sixth form we would be encouraged to read some of the popular mathematical publications in which we would find terms like 'heuristic proof', 'inductive proof', 'iterative proof', and 'illustrative proof'. Sometimes an author would write: 'proceeding formally we can show that . . .', following which a required result would be obtained by apparently flawless techniques which the more naïve reader would accept as a perfectly valid proof.

In this chapter we will study briefly the various types of accepted proofs, consider some of the errors of omission or commission in incomplete proofs, and perhaps clarify some of the misconceptions about what constitutes a proof. There are, however, certain connected points which need to be discussed first.

1. Proof and truth.

2. Inductive proof.

3. Illustrative proofs.

4. Complete and consistent axioms.

5. Lemma and corollary.

6. Necessary and sufficient.

7. The vicious circle.

1. PROOF AND TRUTH

A mathematician does not distinguish between 'proof' and 'truth'. When he says that a certain proposition is 'true', he means that he can justify the result by valid deductive reasoning. The 'absolute truth' of any proposition depends on the truth of the initial statement as well as on the process of logical reasoning, and so we cannot arrive by purely logical arguments at anything we can regard as 'absolutely true'. In other words we regard 'truth' simply as one of the two possible truth values of a proposition—a proposition is either TRUE or FALSE.

2. INDUCTIVE PROOF

The term 'inductive proof' is very frequently misused. When this is used to mean 'proof by mathematical induction', the writer is being very slipshod and misleading. Otherwise the phrase 'inductive proof' is a self-contradiction, because as explained in Chapter 3 the process of induction is the use of plausible or inductive reasoning to arrive at a conjecture which, after adequate confirmation, we regard as a 'probable conjecture'. This conjecture may by a true one, but the inductive procedure by which it was discovered *does not as yet* constitute a mathematical proof. Sometimes a valid proof may be given by a clarification of the inductive procedure used, with careful attention to rigour and precision in statement. Most authors use the term 'heuristic proof' in this sense, i.e., a proof in which

the procedure used can be refined to produce a valid proof. We will therefore disregard the term 'inductive proof'.

3. ILLUSTRATIVE PROOFS

Illustrative proofs are usually 'visible' or geometrical proofs from a figure. Most of us are quite content prior to going up to a university to accept a proof if we can 'see' it in a diagram, and we are then surprised to find that such proofs are not accepted by our new tutors. The evidence that we can tell that something happens in a particular way by looking at a diagram is not admissible in a valid mathematical proof. We are not allowed to say 'it's so' because we can see it in the diagram. If we set out to prove something in mathematics we must prove it by valid mathematical reasoning. We may see in our diagram that a line and a curve intersect, but to prove this simple fact will need the postulate of continuity. In diagrams curves can look continuous and yet possess some peculiar properties.

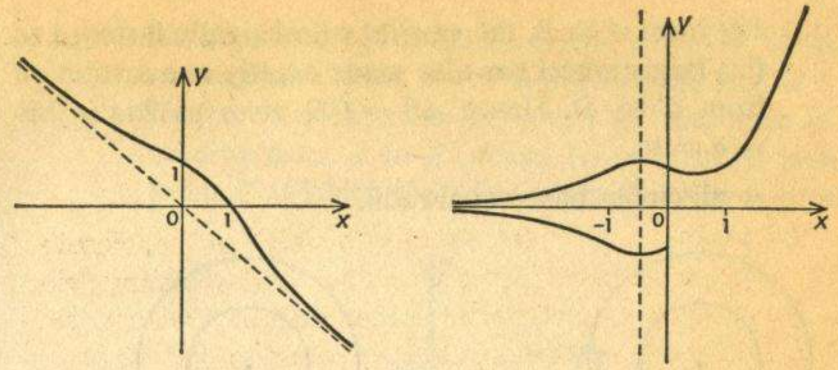
EXAMPLE 1

(*a*) Consider the graph given by $x^3+y^3 = 1$.

Investigation

By Fermat's theorem we know that $x^3+y^3 = z^3$ has no integer solutions. Hence $x^3+y^3 = 1$ will possess no rational solution other than (1,0) and (0,1). Thus the continuous curve defined by this function threads its way through the everywhere dense field of rational points (a,b) without passing through any one of them except (1,0) and (0,1), as shown in figure 4.1(*a*).

(*b*) Consider the graph of $y = x^x$ and its intersection with the straight line $2x+1 = 0$.



$x^3+y^3=1$ (*a*) $y=x^x$ (*b*)

Fig. 4.1

Investigation

For $x > 0$ the value of y is always real and the graph is a continuous curve. For $x < 0$ the value of y may not always exist. If $x = -(p/q)$, a fraction in its lowest terms, we find that $y > 0$ when p is even; $y < 0$ when p is odd and q is odd; y is imaginary when p is odd and q is even. The graph consists of two branches, dense everywhere, as shown in figure 4.1(*b*).

The line $2x+1 = 0$ when drawn in the figure clearly crosses both of these branches. But, when $x = -\frac{1}{2}$ we get $y = \sqrt{-2}$ which is imaginary. Hence the line crosses the curve twice without intersecting it.

EXAMPLE 2

Show that all circles have equal radii (figure 4.2).

Solution

Consider two circular wheels of unequal radii a and b, and fasten the smaller one coaxially to the larger wheel. When the larger wheel has rolled exactly one revolution without slipping along the straight line

PQ from A to B, the smaller wheel rigidly fastened to this larger wheel has also made exactly one revolution from C to D. Hence $AB = CD$ gives us $2\pi a = 2\pi b$ $\Rightarrow a = b$.

$\Rightarrow$ all circles have equal radii.

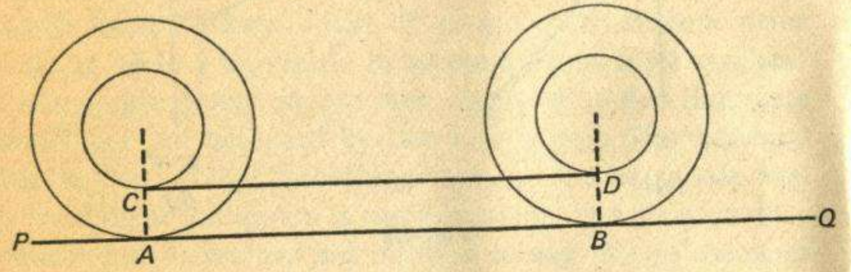


Fig. 4.2

The proposition seems correct visually but our common sense rejects the result. The explanation of this puzzling paradox dates back to the seventeenth century and is closely connected with the strange but true statement that 'no matter how fast a train is moving forwards, certain parts of the train are moving backwards'.

We hope that readers now realize why diagrams may be used to *illustrate* but not to *prove* anything, unless information used from 'seeing' it in the diagram can be established elsewhere in the proof. The visual proof in our next example of a well-known theorem would be quite acceptable in sixth form teaching but would be regarded only as an illustration later on.

EXAMPLE 3

If a real function is continuous in a closed interval $[a,b]$ and if the function assumes opposite signs when $x = a$ and $x = b$, then there exists at least one point c in (a,b) at which the function vanishes.

Proof: The graph of the continuous function $f(x)$ must

be of the general shape shown in figure 4.3. Since it is above the x-axis at $x = a$ and below the axis at $x = b$, the curve must cross the x-axis at least once in proceeding from A to B. Hence there exists at least one point C at which the value of the function is zero.

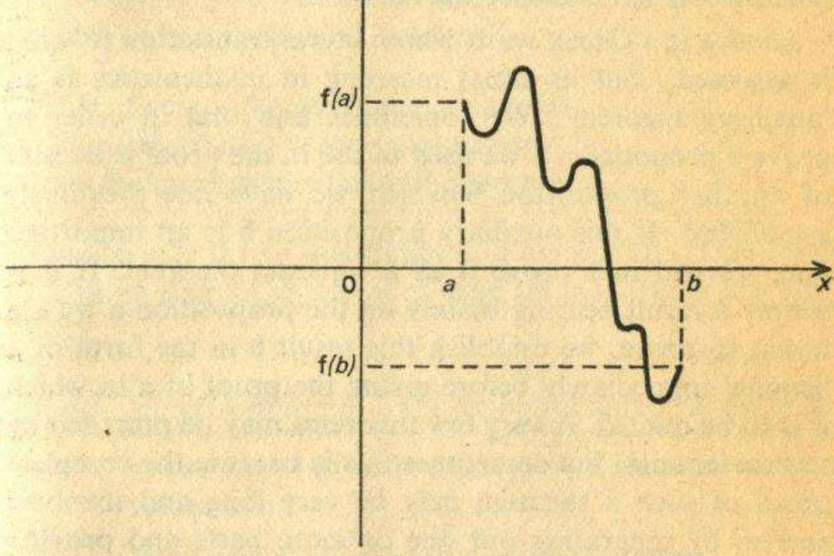


Fig. 4.3

4. COMPLETE AND CONSISTENT AXIOMS

One of the main reasons why visual proofs are regarded purely as illustrations is the fact that Euclid's axioms are incomplete and need to be supplemented by the *axioms of incidence* (i.e. that, when lines are drawn in a definite manner, they intersect in a definite part of the plane). It is easy to show that the breaking of the axioms of incidence gives rise to paradoxes such as the well-known proof that every triangle is isosceles. This difficulty may disappear gradually because of the rapid expansions in the teaching in schools of the so-called 'modern mathematics'. This will lead to a better understanding that a set of axioms must be both 'complete' and 'consistent'—*complete* in that sufficient

axioms have been given for the proving of all propositions deducible in that system, and *consistent* in that one cannot prove that both a proposition and its contradiction are true.

5. LEMMA AND COROLLARY

Lemma is a Greek word whose literal translation is 'what is assumed', but its usual meaning in mathematics is an 'auxiliary theorem'. We sometimes find that in order to prove a proposition **a** we need to use in the proof the result of another proposition **b** which we have not previously established. If this auxiliary proposition **b** is an important one, we will first prove it as a separate theorem. If it is simply a result bearing mainly on the proposition **a** we are trying to prove, we establish this result **b** in the form of a 'lemma' immediately before giving the proof of **a** in which it is to be quoted. A very few theorems may be preceded by several lemmas, but sometimes this is because the complete proof of such a theorem may be very long and involved and so by separating out one or more parts and proving them as lemmas, we shorten the proof to a reasonable length. Some propositions which originated as lemmas became important in their own right, such as Jordan's lemma in complex variable theory.

A *corollary* is a Latin word whose literal translation would be 'gratuity', but its meaning in mathematics is a proposition **c** whose truth follows *immediately* by using another proposition **a** which has just been established. Sometimes several corollaries follow from the proposition **a** and often the argument carries on from one corollary to the next. In other words, either $\mathbf{a} \Rightarrow \mathbf{c}_1$, and $\mathbf{a} \Rightarrow \mathbf{c}_2$, or $\mathbf{a} \Rightarrow \mathbf{c}_1 \Rightarrow \mathbf{c}_2$.

EXAMPLE 4

Proposition **a**.

There exists a unique polynomial in x, of degree $\leqslant n$, which takes on $(n+1)$ given values when x assumes $(n+1)$ *distinct* given values.

Corollary 1: If a quadratic in x vanishes for three distinct values of x, it is the null quadratic.

Proof: From **a** the quadratic is unique.

But $0x^2+0x+0$ is such a quadratic

$\Rightarrow$ the quadratic is the null quadratic.

Corollary 2: If two quadratics in x assume the same values for three distinct values of x, then the quadratics are identical.

Proof: let $ax^2+bx+c = \alpha x^2+\beta x+\gamma$ for three values of x.

$\Rightarrow (a-\alpha)x^2+(b-\beta)x+(c-\gamma) = 0$ for three distinct values of x

$\Rightarrow a-\alpha = 0,\ b-\beta = 0,\ c-\gamma = 0$ from $\mathbf{c}_1$.

Hence $a = \alpha,\ b = \beta,\ c = \gamma \Rightarrow$ quadratics are identical.

6. NECESSARY AND SUFFICIENT

Problems where it is required to find 'necessary and sufficient' conditions or to prove that certain conditions are 'necessary and sufficient' often cause difficulty. This arises because a *necessary* condition, a *sufficient* condition, and a *necessary and sufficient* condition are usually quite different. We can define them as follows:

Definition

If a proposition **p** implies a proposition **q**, we write $\mathbf{p} \Rightarrow \mathbf{q}$ and say that **q** is a *necessary* condition for **p**, and

that **p** is a *sufficient* condition for **q** (i.e. a necessary condition for **p** to be true is that **q** is true, while a sufficient condition for **q** to be true is that **p** is true)

For example, $n \equiv 0 \pmod 9 \Rightarrow n \equiv 0 \pmod 3$.

Here n divisible by 3 is a *necessary* condition for n divisible by 9, while n divisible by 9 is a *sufficient* condition for n divisible by 3.

If we combine these two conditions together we can formulate a necessary and sufficient condition. Thus **q** will be a necessary and sufficient condition for **p** if $\mathbf{p} \Rightarrow \mathbf{q}$ and $\mathbf{q} \Rightarrow \mathbf{p}$, which we write as $\mathbf{p} \Leftrightarrow \mathbf{q}$. For example, $\sin(n\pi)/9 = 0$ is a *necessary and sufficient* condition for $n = 0 \pmod 9$

Hence $$\sin\frac{n\pi}{9} = 0 \Leftrightarrow n \equiv 0 \pmod 9.$$

There are many necessary conditions, many sufficient conditions, but essentially only one set of necessary and sufficient conditions. This set may be stated in several different ways but these will be logically equivalent.

i.e. if $\mathbf{p} \Leftrightarrow \mathbf{q}$ and $\mathbf{p} \Leftrightarrow \mathbf{r}$, then $\mathbf{q} \Leftrightarrow \mathbf{r}$.

Example 5

Necessary and sufficient conditions for a triangle to be right angled can be stated as *any one* of the following:

(*a*) Two of the angles are complementary.

(*b*) The square on one side equals the sum of the squares on the other two sides.

(*c*) The circle on the largest side as diameter will pass through the remaining vertex.

There are many other logically equivalent conditions and readers will recognize the angle-sum, Pythagoras, and the angle in a semicircle theorems of the traditional geometry course.

In the 'to prove' type of question you are given only one such set of conditions to establish as necessary and sufficient conditions. The realization that any logically equivalent set will do may be helpful, i.e. $\mathbf{p} \Leftrightarrow \mathbf{r}$ may be easier to prove than $\mathbf{p} \Leftrightarrow \mathbf{q}$, but the proof will be valid if **q** and **r** are themselves logically equivalent. Remember also that the phrase 'if and only if' (often written 'iff') is another way of writing 'necessary and sufficient'.

Example 6

Prove that a polynomial $f(x)$ is exactly divisible by $(x-2)^2$ if and only if $f(x)$ and $f'(x)$ are each divisible by $(x-2)$.

Necessity

Suppose $f(x) = (x-2)^2Q(x)$, where $Q(x)$ is another polynomial.

$$\begin{aligned} f'(x) &= 2(x-2)Q(x)+(x-2)^2Q'(x) \\ &= (x-2)[2Q(x)+(x-2)Q'(x)], \end{aligned}$$

hence $f'(x)$ is divisible by $(x-2)$.

Sufficiency

Suppose $f(x) \equiv (x-2)Q(x)$ and $f'(x) \equiv (x-2)P(x)$.

$$f'(x) = Q(x)+(x-2)Q'(x) \equiv (x-2)P(x),$$

$$Q(x) \equiv (x-2)[P(x)-Q'(x)],$$

$$f(x) = (x-2)Q(x) = (x-2)^2[P(x)-Q'(x)],$$

hence $f(x)$ is divisible by $(x-2)^2$.

7. *THE VICIOUS CIRCLE*

We sometimes find that a given proposition **a** can be proved quite easily by using some other proposition **b**. This

brings up a very important point—what propositions can we assume to be valid for use in proving a given proposition **a**? Great care is needed to ensure that **b** does not come later in the development of that particular context than **a** does. Even when we can show that **a** and **b** are logically equivalent, we may be 'begging the question', because each of them may have been proved in a similar manner from a common source **k**. Thus there are two types of questions where the use of proposition **b** to prove proposition **a** does not give a valid proof.

(*a*) $\mathbf{a} \Rightarrow \mathbf{b}$ or $\mathbf{a} \Rightarrow \mathbf{k} \Rightarrow \mathbf{b}$.

Clearly we cannot use **b** to prove **a** or else we are arguing in a circle because **a** was used to prove **b**. This kind of error is called a 'vicious circle' and occurred frequently in dictionaries of the last century. Suppose we wished to know what an *ember* was and looked up the definition. We probably found something like:

ember—a live cinder,

cinder—a dead ember.

EXAMPLE 7

(i) Prove $a^2 = b^2+c^2$ for a triangle right-angled at A.

Proof: The cosine rule for triangle ABC is

$$a^2 = b^2+c^2-2bc\cos A$$

since $A = 90°$, $\cos A = 0$

$$\Rightarrow a^2 = b^2+c^2 \text{ when } A = 90°.$$

(ii) If a function $f(x)$ is continuous in closed interval $[a,b]$ and the derivative $f'(x)$ exists in open interval (a,b), prove that $f(a) = f(b)$ will imply the existence of at least one point c in (a,b) at which $f'(c)$ vanishes.

Proof: When $f(x)$ is continuous in $[a,b]$ and $f'(x)$ exists in (a,b), we have an important result called the 'Mean Value Theorem' that there exists a point c in the interval (a,b) at which

$$f(b)-f(a) = (b-a)f'(c).$$

We are given $f(b) = f(a) \Rightarrow (b-a)f'(c) = 0$,

hence $f'(c) = 0$, since $\mathbf{b} \neq \mathbf{a}$.

Each of these proofs is worthless, because the proposition **b** used to establish the result is itself proved by using the proposition **a** as given.

(*b*) $\mathbf{a} \Leftrightarrow \mathbf{b}$ where $\mathbf{k} \Rightarrow \mathbf{a}$ and $\mathbf{k} \Rightarrow \mathbf{b}$.

Here again we cannot use **b** to prove **a** without being sure that $\mathbf{k} \Rightarrow \mathbf{b}$ is not the main point of the question. To prove **a** and **b** are logically equivalent may be only a sophisticated way of saying that a certain proposition can be put in form **a** or in form **b**.

EXAMPLE 8

Show that

$$\int_0^\infty e^{-x^2}dx = \tfrac{1}{2}\sqrt{\pi}.$$

Solution

Put $t = x^2 \Rightarrow dt = 2xdx \Rightarrow dx = dt/2t^{\frac{1}{2}}$, since $x > 0$.

$$\int_0^\infty e^{-x^2}dx = \tfrac{1}{2}\int_0^\infty e^{-t}t^{-\frac{1}{2}}dt = \tfrac{1}{2}\Gamma(\tfrac{1}{2}) = \tfrac{1}{2}\sqrt{\pi}$$

since $\Gamma(\frac{1}{2}) = \sqrt{\pi}$ is a well-known result.

This proof would only be accepted if we are *allowed* to assume the result $\Gamma(\frac{1}{2}) = \sqrt{\frac{1}{2}}$. This result is logically equivalent to the integral required: either can be proved by a

similar process, such as double integration, but we cannot omit this process entirely and deduce each from the other.

We will now return to types of proofs. These will be considered initially under three headings—*direct*, *indirect*, and *incomplete* proofs. Figure 4.4 illustrates the main types of proofs that will be considered and illustrative examples will be given wherever possible.

DIRECT PROOFS

1. METHOD OF EXHAUSTION

The most obvious method of direct proof is that in which you examine *every possible* case and show that the proposition holds in *each* case. This kind of proof is called 'proof by exhaustion'. To examine every possible case is clearly impossible if the set S of possible cases is infinite. Even if S is finite, the enumeration of all members of S and their subsequent examination in each case may be very tedious and lengthy. This examination of each member of S is, however, very useful when the proposition is not true, for all we have to do then is to find just *one* case in which the proposition does not hold. We call such a case a 'counter-example'. If we denote a proposition **p** by $\mathbf{p}_x$ (where x is a variable involved and we know the set S containing all x), then we can represent the above statement by the symbolic statements:

$\mathbf{p}_x$ true for all $x \in S$, proved by exhaustion over S,

$\mathbf{p}_x$ false, if $\sim\mathbf{p}_x$ true for just one $x \in S$.

Even when the set S is not very large, care must be taken to make certain that *every possible* case has been considered. It may happen that despite the incomplete exhaustion the proposition is true, but as shown in the last chapter the proof is not valid until S has been exhausted. There are,

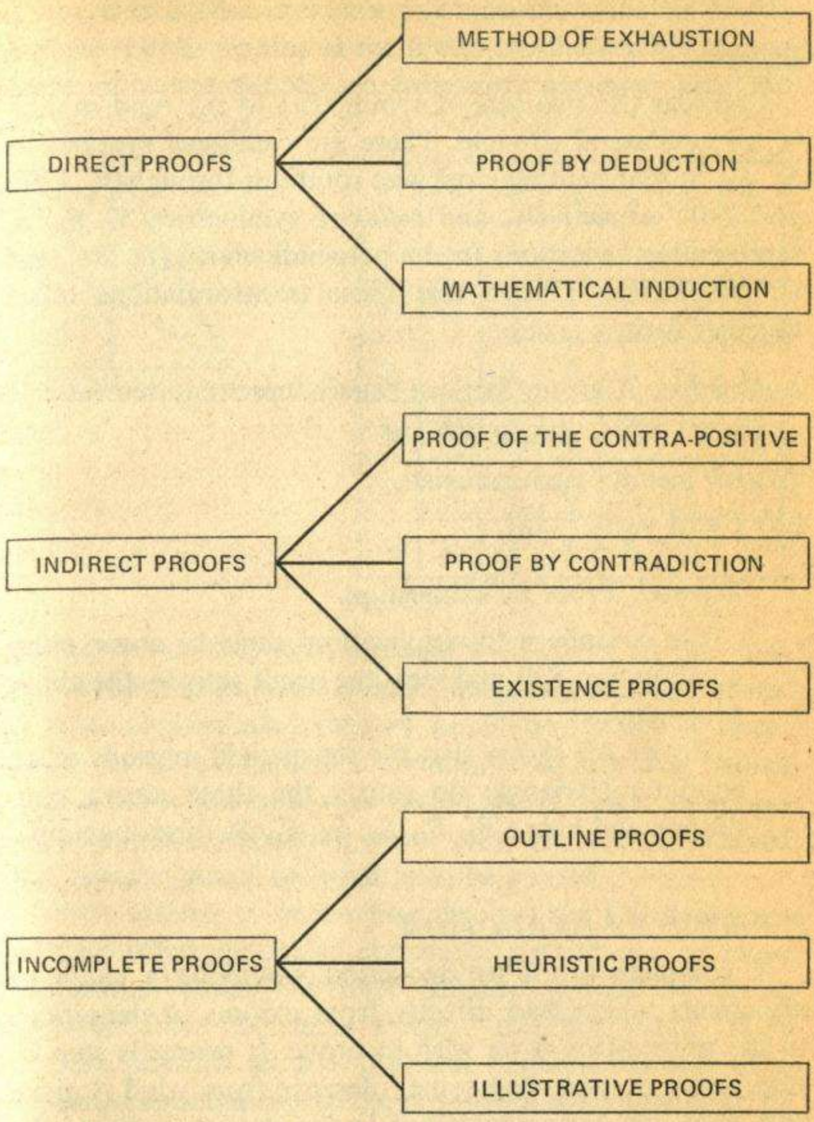


Fig. 4.4

however, certain topics such as *groups* in which proofs by exhaustion are used.

Example 9

Consider the two sets of symmetries in the rigid motion of an equilateral triangle. There are *rotational* symmetries S_1, S_2, S_3 representing clockwise rotations through 0°, 120°, and 240° respectively, and *reflexive* symmetries S_4, S_5, S_6 representing reflections in the perpendiculars AD, BE, and CF respectively. Prove that these transformations taken together form a group.

Knowledge: A group M must satisfy three requirements:

(i) $S, T \in \mathrm{M} \Rightarrow$ unique $ST \in \mathrm{M}$.

(ii) An identity element exists.

(iii) $T \in \mathrm{M} \Rightarrow T^{-1} \in \mathrm{M}$.

Method: Proof by exhaustion.

We compile a 'multiplication' table to cover every combination S_rS_t and see if the result satisfies the above properties.

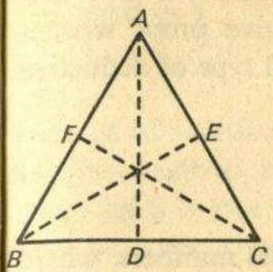
Figure 4.5 shows that the set of rigid motions of an equilateral triangle do satisfy the three above conditions and hence S_1, S_2, S_3, S_4, S_5, S_6 form a group.

2. *PROOF BY DEDUCTION*

A complete proof by deduction would be a chain of arguments which lead directly from axioms or definitions to the proposition $\mathbf{p}$ we wish to prove. It proceeds step by step, using at each step some inference from what is given and what has been established in the preceding steps. We derive $\mathbf{p}$ from a sequence of propositions $\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n$ satisfying the chain:

$$\mathbf{p}_1 \Rightarrow \mathbf{p}_2 \Rightarrow \mathbf{p}_3 \Rightarrow \ldots \Rightarrow \mathbf{p}_n \equiv \mathbf{p}.$$

It is, however, usually very lengthy to deduce propositions from a set of complete self-consistent axioms and definitions. In fact, it is seldom desirable and sometimes impracticable to prove fully every step in the deductive chain given above. Hence deductive proofs are frequently abridged, and the



	S_1	S_2	S_3	S_4	S_5	S_6
S_1	S_1	S_2	S_3	S_4	S_5	S_6
S_2	S_2	S_3	S_1	S_6	S_4	S_5
S_3	S_3	S_1	S_2	S_5	S_6	S_4
S_4	S_4	S_5	S_6	S_1	S_2	S_3
S_5	S_5	S_6	S_4	S_3	S_1	S_2
S_6	S_6	S_4	S_5	S_2	S_3	S_1

Fig. 4.5

extent to which they are shortened will depend on the standard of the readers. Thus a better definition of such a proof for more sophisticated readers would be—'a set of steps, intelligible to the readers for whom it is intended, which point to the existence of a deductive proof'. But, no matter how abridged a deduction proof may be, it must proceed from 'what is given' to 'what is to be proved'.

Hence, instead of proceeding through the full deductive chain we often choose as our starting point one or more propositions which we have already proved by direct deduction from the same axioms and definitions. It is therefore possible to give different deductive proofs of a proposition $\mathbf{p}$ according to the propositions we assume to be true without preliminary proof. If we have previously in the development of a subject, proved $\mathbf{p}_1 \Rightarrow \mathbf{q}_1$, we replace the chain $\mathbf{p}_1 \Rightarrow \mathbf{p}_2 \Rightarrow \mathbf{p}_3 \Rightarrow \ldots \Rightarrow \mathbf{p}_n \equiv \mathbf{p}$ by a new chain $\mathbf{q}_1 \Rightarrow \mathbf{q}_2 \Rightarrow \ldots \Rightarrow \mathbf{q}_m \equiv \mathbf{p}$.

Thus a clear understanding of which earlier propositions may be assumed without proof is essential in giving deductive proofs. As mentioned earlier in discussing the 'vicious circle' error it is very important to check that the proposition $\mathbf{q}_1$ does not come later in the development of the context than the given proposition $\mathbf{p}$ does. We give two examples, the first illustrating a deductive proof written out in full and the second the more usual type of deductive proof.

Example 10

A binary operation on the set $\mathbf{R}$ of real numbers, where addition and subtraction have their usual meaning, is defined by the relation

$$x \circ y = x+y+xy; \qquad x, y \in \mathbf{R}.$$

Prove there is a unique identity element $e \in \mathbf{R}$.

Proof

$x \circ y = x+y+xy; \quad x, y \in \mathrm{S}$	definition of $\circ$
$y \circ x = y+x+yx$	definition of $\circ$
$x+y+xy = y+x+yx$	commutivity of addition and multiplication
$x \circ y = y \circ x$	equivalence
Thus the operation is commutative	definition
Suppose e and e_1 are two identity elements in S	
$x \circ e = x$ and $x \circ e_1 = x$	definition of identity element
Hence $e_1 \circ e = e_1$ and $e \circ e_1 = e$	rule of substitution
$e = e \circ e_1 = e_1 \circ e = e$	commutivity proved
Hence there is at most one identity element	
$0 \circ y = 0+y+0y = 0+y+0 = y$	0 is unit element in addition and the zero element in multiplication.
Hence 0 is the unique identity element in S	

The rule of substitution is the rule of inference which states that 'whatever is true of an arbitrary element of a *non-empty* set S is true of a particular element of that set'. Care must be taken to see that set S is non-empty; otherwise a fallacy may arise as is seen in Example 8 on mathematical induction.

Example 11

If the coefficients are real with at least one a_r non-zero, prove that the equation

$$x^6+a_1x^3+a_2x^2+a_3x+a_4 = 0$$

has at least two complex roots.

Investigation

We do not think of attempting to prove this in full from first principles, but we try to think of any relevant knowledge that we have established in polynomial theory. We should have little difficulty in giving a proof as follows:

Proof: Let the roots be $r_1, r_2, \ldots, r_6$. (i)

From elementary theory of equations

$$\sum_1^6 r = 0 \quad \text{and} \quad \sum_{i<j}^6 r_i r_j = 0 \qquad \text{(ii)}$$

Hence $\qquad \sum_1^6 r^2 = (\sum r)^2 - 2\sum r_i r_j = 0 \qquad$ (iii)

$\Rightarrow$ at least one of the roots is complex. (iv)

But complex zeros of a polynomial with real coefficients occur in conjugate pairs, hence

there must be at least two complex roots. (v)

Notes

(i) This line uses the established proposition $\mathbf{q}_1$ that a polynomial of degree n has n zeros over the complex field.

(ii) This line uses established proposition $\mathbf{q}_2$ that gives the connections between the coefficients of a polynomial equation and certain symmetric functions of the roots.

(iii) This line is an algebraic identity.

(iv) This is all that is implied by the previous step, because we see that

$$2^2+2^2+2^2+2^2+3^2+(5i)^2 = 0.$$

(v) This line quotes an extablished proposition $\mathbf{q}_3$.

Hence our proof contains two deductive steps and three established propositions from polynomial theory.

One last point should be noted. When we show by valid mathematical reasoning that proposition **p** implies proposition **q**, we must realize that there are three possibilities:

1. If **p** is true, then **q** is true.
2. If **q** is false, then **p** is false.
3. If **p** is false, then **q** may be true or false.

EXAMPLE 11A

The statement

$$x < 1 \Rightarrow x < 2$$

is true for all x in **R**. This includes all the following statements, each of which is true:

$0 < 1 \Rightarrow 0 < 2$, **p** true and **q** true
$1 < 1 \Rightarrow 1 < 2$, **p** false and **q** true
$2 < 1 \Rightarrow 8 < 2$, **p** false and **q** false.

MATHEMATICAL INDUCTION

Mathematical induction is the name given to a method of proof applicable to a proposition in which a variable ranges over the set $\mathbf{Z}^+$ or else the set $\mathbf{Z}^+_{n \geqslant n_0}$. We wish to prove a proposition for an infinite number of integral values of this variable and the whole basis of our proof depends on the following axiom called the 'axiom of mathematical induction'.

Axiom: If a set $S \subset \mathbf{Z}^+$ and

(i) The integer $1 \in S$,
(ii) The integer $k+1 \in S$ whenever $k \in S$,
then $S = \mathbf{Z}^+$.

The extension to the more general induction axiom will be

(i) The integer $n_0 \in S$.
(ii) The integer $k+1 \in S$ whenever $k \in S$,
then $S = \mathbf{Z}^+_{n \geqslant n_0}$.

The method of establishing that a proposition $\mathbf{p}_n$ is true for all $n \in \mathbf{Z}^+$ has two parts:

(i) Show $\mathbf{p}_n$ is true for $n = 1$.
(ii) Show that $\mathbf{p}_k$ true $\Rightarrow \mathbf{p}_{k+1}$ true, for $k \in S$.

Then the *induction axiom* establishes the result for all $n \in \mathbf{Z}^+$.

In practical examples the essential first part that $\mathbf{p}_1$ is true is always elementary, but the second part may involve considerable manipulation. There are two very common types of manipulative working:

(*a*) Those in which $\mathbf{p}_{k+1} = \mathrm{T}(\mathbf{p}_k)$ where T is some definite operation performed on $\mathbf{p}_k$ to get $\mathbf{p}_{k+1}$. We perform this operation on the result assumed to be true for $\mathbf{p}_k$ and show that it can be evaluated to $\mathbf{p}_{k+1}$.

(*b*) Those in which we use the identity $\mathbf{p}_{k+1} \equiv \mathbf{p}_k + (\mathbf{p}_{k+1} - \mathbf{p}_k)$. We evaluate the difference $\mathbf{p}_{k+1} - \mathbf{p}_k$ directly and then show that the result of adding this difference to the assumed result for $\mathbf{p}_k$ leads to $\mathbf{p}_{k+1}$.

The first of the following examples illustrates the vital necessity of proving the first part of the method, i.e. proving that S is not the empty set.

EXAMPLE 12

(*a*) Prove that $\sum\limits_1^n 2^r = 2^{n+1}$ for all $n \in \mathbf{Z}^+$.

Assume result true for some $n = k \in \mathbf{Z}^+ \Rightarrow \sum\limits_1^k 2^r = 2^{k+1} = \mathrm{f}(k)$

$$\begin{aligned}\mathbf{p}_{k+1} &= \sum_1^{k+1} 2^r = \sum_1^k 2^r + 2^{k+1} = 2^{k+1} + 2^{k+1} = 2^{k+2} \\ &= \mathrm{f}(k+1),\end{aligned}$$

hence $\qquad \mathbf{p}_k \Rightarrow \mathbf{p}_{k+1}$.

But the result is not true because there is no value of $k \in \mathbf{Z}^+$ for which it is valid.

(*b*) Prove that $n^2 > 2n+100$ for $n \in \mathbf{Z}^+$.

Assume true for some $k \in \mathbf{Z}^+ \Rightarrow k^2 > 2k+100 = \mathrm{f}(k)$

$$\begin{aligned}(k+1)^2 = k^2+(2k+1) &> 2k+100+(2k+1) \\ &= 2(k+1)+100+(2k-1) \\ &\Rightarrow (k+1)^2 > 2(k+1)+100 \text{ for } k \geqslant 1.\end{aligned}$$

Again the result fails to be true for all $n \in \mathbf{Z}^+$, because when $n = 1$ we have $1 > 102$, but this time the set is not empty. When $n = 12$ we have $\mathbf{p}_k = 12^2 = 144$ and $2k+100 = 124$.

Hence $\qquad \mathbf{p}_k > 2k+100$ for $k = 12$

$\Rightarrow$ proposition is true for $n \geqslant 12$ by mathematical induction.

EXAMPLE 13

Prove that

$$\mathrm{D}^n(\mathrm{e}^{3x} \sin 4x) = 5^n \mathrm{e}^{3x} \sin (4x+na) \text{ where } \tan a = 4/3.$$

Proof: $\mathrm{Df} = \mathrm{e}^{3x}(3 \sin 4x + 4 \cos 4x)$

$$\begin{aligned}&= 5\mathrm{e}^{3x}(\tfrac{3}{5} \sin 4x + \tfrac{4}{5} \cos 4x) \\ &= 5\mathrm{e}^{3x} \sin (4x+a) \Rightarrow \text{result is true for } n = 1.\end{aligned}$$

Assume true for some $k \in \mathbf{Z}^+ \Rightarrow \mathrm{D}^k\mathrm{f} = 5^k \sin (4x+ka)\mathrm{e}^{3x}$

$$\begin{aligned}\mathrm{D}^{k+1}\mathrm{f} = \mathrm{D}\{\mathrm{D}^k\} &= 5^k\mathrm{e}^{3x}[3 \sin (4x+ka)+4 \cos (4x+ka)] \\ &= 5^{k+1}\mathrm{e}^{3x} \sin \{4x+(k+1)a\}.\end{aligned}$$

Hence k true $\Rightarrow k+1$ true,

and hence result is established for all $n \in \mathbf{Z}^+$ by mathematical induction.

Students sometimes fear that even if they should begin with the *wrong* conjecture, this method will still prove it

to be correct. They need not worry because the wrong conjecture either leads to the empty set or else the step k true $\Rightarrow$ $(k+1)$ true fails. The case of the empty set is seen in Example 12(a), and the following example shows the failure otherwise.

EXAMPLE 14

Show that

$$1+2+3+\ldots+n = n^2$$

Proof: When $n = 1$ we have $S(1) = 1$ and $f(1) = 1^2 \Rightarrow$ result true for $n = 1$.

Assume $S_k = k^2$ for some $k \in \mathbf{Z}^+$.

$$S_{k+1} = S_k+(k+1) = k^2+k+1 \neq (k+1)^2 \text{ as required.}$$

Hence $\quad S_k$ true $\not\Rightarrow S_{k+1}$ true

$\Rightarrow$ conjecture is false.

INDIRECT PROOFS

There are three main types of indirect proofs—*proof of the contra-positive*, *proof by contradiction*, and *existence proofs*. Many other methods of proof are essentially variations of these three methods. For example, Fermat's method of infinite descent is a variation of the proof by contradiction. We will now consider these three types of proof in detail.

PROOF OF THE CONTRA-POSITIVE

This method of proof uses the principle that the implication $\mathbf{p} \Rightarrow \mathbf{q}$ is logically equivalent to the implication $\sim\mathbf{q} \Rightarrow \sim\mathbf{p}$. This implication is called the contra-positive of $\mathbf{p} \Rightarrow \mathbf{q}$ and must not be confused with the converse of $\mathbf{p} \Rightarrow \mathbf{q}$

(which is $\mathbf{q} \Rightarrow \mathbf{p}$). The method is evident from the following example.

EXAMPLE 15

Show that for all $n \in \mathbf{Z}^+$, n is not divisible by k if n^2 is not divisible by k.

Proposition: n^2 not divisible by $k \Rightarrow n$ not divisible by k.

Contra-positive: n divisible by $k \Rightarrow n^2$ divisible by k.

If n divisible by k, $n = mk$

$$n^2 = m^2k^2 = k(m^2k) \text{ is also divisible by } k,$$

hence contra-positive is true $\Rightarrow$ original proposition is true.

It is true that this method of proof is seldom used in practical examples and yet it gives a simple proof of some of the elementary propositions in number theory and analysis.

EXAMPLE 16

Every bounded non-empty set of natural numbers contains its upper bound.

Proposition: Bounded set S of natural numbers $\Rightarrow$ upper bound $M \in$ S.

Contra-positive: Set S of natural numbers contains no upper bound $\Rightarrow$ S is unbounded.

Since S is non-empty, it contains a natural number K.

Since K is not the upper bound, S contains natural number $K_1 > K$. Repetition shows that S contains a sequence of natural numbers $\{K_r\}$ such that

$$K < K_1 < K_2 < K_3 < \ldots$$

By the axiom of induction the set S will be unbounded. Hence contra-positive is true $\Rightarrow$ proposition is true.

Sometimes the example may look very trivial and yet it may be quite an important result when put in another logically equivalent way. For example, consider the proposition:

If x is a real number, then $x < \varepsilon$ for all $\varepsilon < 0 \Rightarrow x \leqslant 0$.

The contra-positive is evidently true by choosing $x = \varepsilon$ and so the given proposition is true.

This looks quite trivial, but what about one logically equivalent form:

$$\bigcup_{n=1}^{\infty}\left[\frac{1}{n},1\right] = \left[0,1\right]?$$

PROOF BY CONTRADICTION

In order to prove a proposition **a** by this method we prove that the proposition $\sim$**a** is *false*. We begin the proof by assuming that $\sim$**a** is *true* and show by mathematical reasoning that this leads to a contradiction. For example, to prove the implication **p** $\Rightarrow$ **q** we assume both **p** and $\sim$**q** are true and deduce from them both a result **c** and its negation $\sim$**c**. If the mathematical reasoning is valid, then the false result must be due to the falsity of our original assumption that **p** $\Rightarrow$ $\sim$**q**, and hence the result **p** $\Rightarrow$ **q** must be true. Sometimes we are able to deduce from **p** $\Rightarrow$ $\sim$**q** a new proposition **c** which contradicts a well-known result in the context of the system in which we are working.

This method of proof is by far the commonest indirect method and, indeed, furnishes the only known method of establishing many important theorems in various branches of mathematics. We originally met this method in our traditional school geometry course, under the name

reductio ad absurdum, as a common method of proving the converse of established theorems. For example, after proving the theorem that 'opposite angles of a cyclic quadrilateral are supplementary', all textbooks prove the converse by assuming that the circle passing through three vertices of a quadrilateral whose opposite angles are supplementary does not pass through the fourth vertex, i.e. that **p** $\Rightarrow$ $\sim$**q**.

It is of interest to note that propositions provable by mathematical induction can also be proved by this method, as illustrated in the second of the following examples.

EXAMPLE 17

The two classic examples using proof by contradiction:

(*a*) The number of primes is infinite.

Proof: Assume the number of primes is not infinite, and so there exists a greatest prime p.

Consider the number $p!+1$.

It has remainder 1 when divided by 2, 3, 4, . . ., p and so is not divisible by any integer $\leqslant p$

$\Rightarrow p!+1$ is either a prime or else is divisible by a prime greater than p, which gives a contradiction (since p assumed the largest prime).

$\Rightarrow$ no greatest prime exists $\Rightarrow$ number of primes is infinite.

(*b*) $\sqrt{2}$ is irrational.

Proof: Let $\sqrt{2} = \frac{a}{b}$ where a, b are rational and have no common divisor.

$$2 = \frac{a^2}{b^2} \Rightarrow a^2 = 2b^2 \Rightarrow a \text{ is even.}$$

Suppose $a = 2m \Rightarrow 4m^2 = 2b^2 \Rightarrow 2m^2 = b^2$.

Hence b is even, which gives a contradiction (since b must be odd when a is even in order that a/b has no common divisor).

$$\text{Hence } \sqrt{2} \neq \frac{a}{b} \text{ for any rational numbers } a, b$$

$$\Rightarrow \sqrt{2} \text{ is irrational.}$$

Example 18

Show that for $n \in \mathbf{Z}^+$, the sum to n terms of series

$$1+3+5+7+\ldots = n^2.$$

Proof: Suppose result is not true for some $k \in \mathbf{Z}^+$.

$$S_k = \sum_1^k (2r-1) \neq k^2.$$

$$S_{k-1} = S_k-(2k-1) \neq k^2-(2k-1) = (k-1)^2.$$

Repetition leads to $S_1 \neq 1^2$, which gives a contradiction since $S_1 = 1$.

$\Rightarrow \sum_1^n (2r-1) = n^2$ is correct for all $n \in \mathbf{Z}^+$ by the axiom of induction.

EXISTENCE PROOFS

There are two distinct types of existence proofs—the *purely existence* proof and the *constructive existence* proof. The ability to prove the existence of objects without actually exhibiting one of them is one of the most distinctive features of mathematics. The constructive procedures that arise in the constructive existential proof are called algorithms and are of great importance. Such proofs are

frequently called 'iterative' proofs and the algorithms are called 'iterative procedures'. A necessary consideration in tackling certain types of problems (such as isoperimetric problems) is to be sure that a solution exists. Mathematicians tried for many years to determine the least area in which to turn a needle end for end, only to discover eventually that there is no least area.

Illustrative example

A given number of radar observation posts are to be disposed on the surface of the earth. What is the best disposal position, where *best* means most widely separated, each from each?

For some given numbers the solution is quite easy:

four posts are best dispersed at the vertices of a regular inscribed tetrahedron.

six posts are best dispersed when 90° apart, such as one at each pole and the other 4 at equally spaced sites on the equator.

What about 5 posts or 7 posts? Analysis shows that there does exist a best dispersion for 7 posts, but that *there is no solution for 5 posts.*

Oddly enough, there is a best dispersion for 8 posts but it is *not* the sites at the 8 vertices of the inscribed cube.

The two types of existence proofs are illustrated by the proofs given for the existence of transcendental numbers by Liouville and Cantor. Liouville exhibited a transcendental number, whereas Cantor proved that a transcendental number was a member of a non-empty set by showing that the set of all algebraic numbers is smaller than the set of all real numbers. The relevant proofs can be found in most textbooks on number theory or analysis. Here is a practical example of Brouwer's *fixed point theorem* proved by both methods.

EXAMPLE 19

Two different-sized circular maps of the same geographical area are placed on a table with the larger map completely over-lapping the smaller one. Show that there exists one point at which a pin can be inserted to pierce the *same* place on each map (figure 4.6).

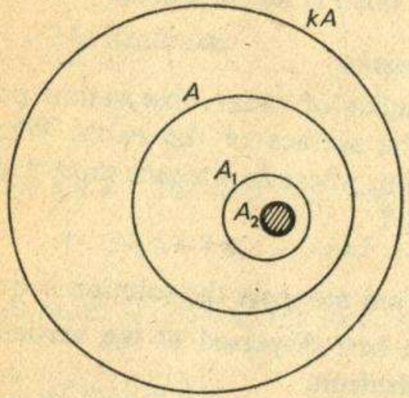


Fig. 4.6

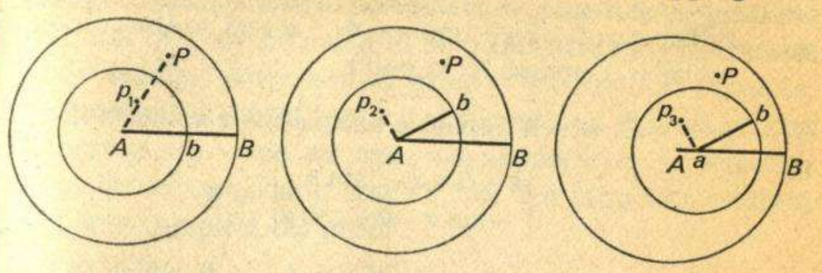
(*a*) Existence proof

Suppose the large map has scale k times the small map which has area A. Consider the region on the large map lying immediately below the small map. This region will be mapped on a region A_1 (of area A/k^2) of the small map. Similarly the region on large map immediately underneath area A_1 will be mapped on a region A_2 (of area A_1/k^2) on the small map. When we continue this sequence of mappings, we have $A_{n-1} \to A_n$ where A_n has area A/k^{2n}. Since $k > 1$, given $\varepsilon > 0$ we can find N so that $A/k^{2n} < \varepsilon$ for all $n > N$.

Hence, as n increases indefinitely, the area A_n will become indefinitely small with limit approaching 0 as n gets large enough. The limit will be a point P, and a pin inserted at this point will pierce the same place on each map.

(*b*) Constructive proof

Suppose A and a are the respective centres of the large and small circle, and AB is the radius along the line of centres Aa (figure 4.7). The final mapping can



(a) Reduction *(b)* Rotation *(c)* Translation

Fig. 4.7

be considered as three successive simple mappings: reduction on scale $k:1$, rotation through α anticlockwise and translation from $A(0,0)$ to $a(c,0)$. Let P have coordinates (x,y) referred to origin A and AB as $0x$ axis.

The reduction maps P onto p, where $x_1 = \frac{1}{k}x$ and $y_1 = \frac{1}{k}y$.

The rotation maps p_1 onto p_2, where

$$x_2 = x_1 \cos\alpha - y_1 \sin\alpha,$$

$$y_2 = x_1 \sin\alpha + y_1 \cos\alpha.$$

The translation maps p_2 onto p_3, where

$$x_3 = x_2 + c,\ y_3 = y_2.$$

Hence the final result is that $P \to p_3$ where

$$x_3 = \frac{1}{k}(x\cos\alpha - y\sin\alpha) + c$$

$$y_3 = \frac{1}{k}(x\sin\alpha + y\cos\alpha)$$

Any fixed point of this mapping must satisfy

$$\begin{aligned} (k-\cos\alpha)x+y\sin\alpha &= kc \\ -x\sin\alpha+(k-\cos\alpha)y &= 0 \end{aligned} \tag{1}$$

Either $c = 0$ which implies $A \equiv a$ and so the common centre is a unique fixed point,

or $c \neq 0$, which implies a unique finite solution iff

$$\begin{vmatrix} k-\cos\alpha & \sin\alpha \\ -\sin\alpha & k-\cos\alpha \end{vmatrix} \neq 0$$

This requires $(k-\cos\alpha)^2+\sin^2\alpha \neq 0$ which is true for all α, since $k > 1$. Hence for all values of c, α, $k(>1)$ there exists a unique point P which is mapped on itself. The coordinates of this point are obtained by solving the simultaneous equations (1).

Note

The constructive type of existence proof usually follows the pattern of the purely existence proof given above. Our problem defines a particular problem space (usually a function space) in which we can determine the effect of a certain mapping T on an element of our space, just as in the problem above we have $\mathrm{T}(P) = p_3$ in the Euclidean space $\mathbf{R}^2$. We then have to show that:

(i) The sequence $\{\mathrm{T}^n(P)\}$ converges to a limit P^*.

(ii) the limit P^* lies in our problem-space.

(iii) P^* satisfies the conditions of the problem.

Hence we deduce P^* exists, and the algorithm $\mathrm{T}(P_n) = P_{n+1}$ furnishes a formula from which we can determine P^* to any required degree of accuracy. Such algorithms are usually called 'iterative' formulae and the constructive-existence type of proof an 'iterative' proof.

INCOMPLETE PROOFS

Incomplete proofs of various types are frequently encountered in elementary textbooks and in books on 'popular' mathematics. Sometimes the authors state that the relevant arguments are incomplete or plausible or heuristic and that they are intended to be illustrative only as distinct from definite mathematical proofs. We should, however, realize what is meant by such statements so that we can be on our guard when we meet arguments like: 'proceeding formally we can show that . . .' or 'an induction method will now complete the proof'.

OUTLINE PROOFS

Here a sketchy proof of a proposition is given with considerable gaps in the arguments quoted. Often such gaps are the omission of much manipulative work and the reader can remedy this by taking out paper and pencil and checking that the results quoted do arise from the arguments in the previous steps. Where the gaps do omit more than manipulative work the proof is probably intended only for those readers who have sufficient background knowledge (of the particular context) to realize what has been omitted. By outline proofs we understand the former type—proofs where the interested reader could fill in the missing gaps and so finally obtain a complete proof.

EXAMPLE 20

(*a*) Show that at least two people in Glasgow have exactly the same number of hairs on their head.

Proof: Use Dirichlet's *pigeonhole principle.*

(*b*) If p is a prime number, prove that $(p-1)!+1$ is divisible by p.

Proof: Under multiplication the residues modulo p form a commutative group.

$$(p-1)! = 1\times2\times3\times \dots \times(p-1).$$

Because of commutivity we can pair off every element with its inverse and we will then be left only with $(p-1)$

$$\Rightarrow (p-1)! \equiv p-1 \pmod p$$

$$\Rightarrow (p-1)!+1 \equiv 0 \pmod p.$$

In (*a*) although only one line is given, any student who is familiar with Dirichlet's pigeonhole principle—'if m objects be distributed in n pigeonholes and if $m > n$ then at least one pigeonhole contains at least two objects'—should be able to finish the solution. It seems reasonable that any person would collapse under the weight of one million hairs (especially those of the modern variety!) and we know that Glasgow has more than one million inhabitants.

In (*b*) only a sophisticated reader with a good background knowledge of groups could properly understand the proof given. The amount of work to be filled in could be very considerable for a reader with only a scanty knowledge of groups. Remember that the binary operation here is defined as:

$$i \circ j = \text{residue } i\times j \pmod p.$$

HEURISTIC PROOFS

The term 'heuristic proof' is used in different senses by different authors and the meaning can vary from a purely *plausible argument* to a reasonable *outline* proof. The purely plausible argument does not in any degree constitute a mathematical proof, but furnishes only a confirmation

that the proposition may be true. When used in the other extreme sense of an outline proof, the argument given may be capable of improvement to a rigorous proof and the term 'heuristic' should not be applied to such arguments. The usual meaning of a heuristic proof is an exercise of reasoning which suggests the correct solution and points the direction in which a rigorous proof may be sought. The usual chain of logical deduction is broken by a step in which a more or less plausible assumption is made. Such a proof has a suggestive value and is only provisional until we can justify rigorously the plausible assumptions made. Heuristic procedures are of frequent occurrence in applied mathematics where we often assume a solution exists or that there exists a mathematical law or function of a certain type which relates to the problem in hand. Once such assumptions are made it is usually quite easy to evaluate the solutions.

The commoner assumptions made in such heuristic proofs have been already discussed under 'assumption' in Chapter 3 and the two exercises in Example 17 illustrate what can happen when you use procedures which you are unable to justify in the particular problem under investigation. It is probably because heuristic arguments frequently arise in investigations by the inductive method that some writers use the misleading term 'inductive proof' where they really mean 'heuristic proof'. Here is one of the well-known heuristic proofs from Euler, the greatest of all the providers of heuristic proofs.

EXAMPLE 21

Show that

$$\sin \pi x = \pi x\left(1-\frac{x^2}{1^2}\right)\left(1-\frac{x^2}{2^2}\right)\left(1-\frac{x^2}{3^2}\right)\dots.$$

Proof: We know that a polynomial f(x) of degree n

with n distinct zeros can be written in either of the forms

$$a_nx^n+a_{n-1}x^{n-1}+\ldots+a_0$$
$$=a_n(x-x_1)(x-x_2)(x-x_3)\ldots(x-x_n)$$
$$=K\left(1-\frac{x}{x_1}\right)\left(1-\frac{x}{x_2}\right)\left(1-\frac{x}{x_3}\right)\ldots\left(1-\frac{x}{x_n}\right)$$

and we see $K=a_0$ by equating the constant term on each side.

Now Euler made the plausible assumption that a similar product decomposition of a non-polynomial function $f(x)$ may be made if we know all the zeros of that function.

He knew that $\sin\pi x$ had zeros $0, \pm1, \pm2, \pm3, \ldots$, and that

$$\sin x = x-\frac{1}{6}x^3+\frac{1}{120}x^5-\ldots.$$

He thus made the plausible suggestion

$$\sin\pi x = Kx\left(1-\frac{x^2}{1^2}\right)\left(1-\frac{x^2}{2^2}\right)\left(1-\frac{x^2}{4^2}\right)\ldots,$$

and on equating coefficients of x on each side he got $K=\pi$. This led to his celebrated formula:

$$\sin\pi x = \pi x\left(1-\frac{x^2}{1^2}\right)\left(1-\frac{x^2}{2^2}\right)\left(1-\frac{x^2}{3^2}\right)\ldots.$$

The master, as usual, had exercised his judgement correctly, because years later, when convergency of series and infinite products had been studied, it was seen that this infinite product converges for all values of x. The danger in such heuristic arguments without their subsequent rigorous justification is seen by taking $f(x)=e^x\sin\pi x$.

$e^x\sin\pi x$ has zero $0, \pm1, \pm2, \pm3, \ldots$, since $e^x \neq 0$.

Exactly the same heuristic argument leads to the result

$$e^x\sin\pi x = \pi x\left(1-\frac{x^2}{1^2}\right)\left(1-\frac{x^2}{2^2}\right)\left(1-\frac{x^2}{3^2}\right)\ldots.$$

The error in the proof is that the new function $e^x\sin\pi x$ does possess another zero, $x=-\infty$. Although the infinite product converges for all finite x, it does not contain all the zeros of $e^x\sin\pi x$, and so the possible analogy with polynomial equations disappears.

ILLUSTRATIVE PROOFS

Illustrative proofs are usually of three types:

(i) *Proof of a particular case.*

(ii) *Proof of analogous cases in another context.*

(iii) *Diagrammatic proofs.*

The proof of a particular case (often a numerical example) is simply a confirmation that the proposition may be true, and is in no sense to be considered as a valid proof of the proposition. It may happen that the method used to prove the particular case can be extended without change of form to the general case, but until this is fully explained at the end of the proof for the particular case the arguments used simply point the way to a valid mathematical proof.

The use of analogous propositions in another context are usually stated to be purely illustrative. It may be that we can show that the new system and the original system have identical structures. Then a proof of such isomorphism may enable the illustrative argument to be made into a valid mathematical proof of the original proposition. But, until the connection between the two structures is clarified, the argument given is again only a confirmation that the original proposition may be true. In other words, illustrative proofs of particular cases or analogous cases are suggestive

and encouraging, but they are not mathematical proofs of the original proposition.

The diagrammatic proofs do raise some difficulties as we have already seen earlier in this chapter when discussing *geometrical proofs*. At pre-university level diagrammatic proofs are usually accepted as valid mathematical proofs. At first-year university level tutors usually take the view that diagrams may be used to *illustrate* but not to *prove* anything. We are not allowed to rely on a diagram for any information not furnished elsewhere in the proof. On the other hand, geometrical proofs of geometrical propositions are quite valid. In other words, diagrammatic proofs are in general regarded simply as pictorial presentations or illustrations, unless the proposition is formulated in a geometrical context. One of the arguments against diagrammatic proofs is that so much extraneous material has to be introduced before the proof becomes intelligible. Such an objection cannot really be answered because, logically, the validity of a proposition is independent of the way in which it is proved.

EXAMPLE 22

If sets A, B satisfy $B \subseteq A$, show that $A-(A-B) = B$.

First proof: Let A be the set of all real numbers and B the set of all rational numbers. Then $A-B$ is the set of all irrational numbers. If we subtract this set from the set of all real numbers, we will be left with the set of all rational numbers

$$\Rightarrow A-(A-B) = B.$$

Second proof: The Venn diagram (figure 4.8) satisfies the condition that $B \subseteq A$. We see from the diagram

that $A-B$ is represented by the shaded area and that, when we subtract this shaded area from A we have the set B remaining

$$\Rightarrow A-(A-B) = B.$$

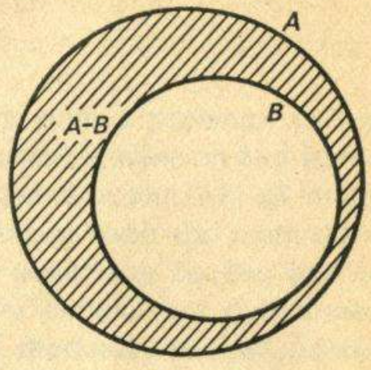


Fig. 4.8

Each proof seems reasonable but neither is acceptable as a valid mathematical proof. The first proof is the proof of a particular case and the second proof requires that we *see* something in a diagram. We would say that the first one is an *illustration* and the second a *pictorial representation*. A valid proof would be something like this:

Third proof: If $B = A$ or $B = \emptyset$, the result is trivially true.

Otherwise,

$$x \in B \Rightarrow x \in A \text{ and } x \in A-B \Rightarrow x \in A-(A-B)$$

Hence $\quad B \subseteq A-(A-B)$.

$$x \in A-(A-B) \Rightarrow x \in A-B$$

and $\quad x \in A \Rightarrow x \in B.$

Hence $\quad A-(A-B) \subseteq B.$

These two results show that $A-(A-B) = B$.

Fourth proof: As an exercise in logic, the statement

$$\sim(\sim B) = B$$

seems to furnish a complete proof.

5. Problems and Extensions

Hitherto we have studied problems that illustrated some particular point under discussion and have not carried out the last step suggested in our over-all strategy—*the review*. In reviewing solutions even the most successful problem-solver may find something he has overlooked. He may find an alternative method that gives a more elegant solution, or a way of shortening part of the solution, or a new approach to the problem that may lead to a further study of some particular mathematical topic. After deciding that the proof is the best we can find, we then consider whether the result proved or the method of proof used can be applied to other related problems.

The application of a general result to numerical cases can hardly be termed an extension of the problem, and, if the inductive method was used in making the original conjecture, such specialization will have already been carried out. What we really need to do is to construct *new* problems related in some way to the problem in hand. We may, for instance, change some of the data in the original hypothesis and then study its effect on our solution. Sometimes the related problems prove inaccessible both in regard to the method used to prove our problem and in the application of the result. Yet by such reviews we gradually develop our ability to solve problems and to discriminate critically between solvable problems and inaccessible problems. It must be admitted that the finding of new interesting problems which become more accessible because of the problem in hand requires experience and luck, but it gives a real

boost to our confidence when we do find and solve such related problems on our own initiative. The desire to explore is a very useful asset in problem-solving and should be very much in evidence at the review stage. We shall now consider some interesting problems, tackling them in the spirit in which the book is written and seeking to find related problems or extensions, wherever possible.

PROBLEM 1. A SHORTEST DISTANCE PROBLEM

Two towns A and B lie on the same side of a straight canal and their main sewage pipes are connected to a common outlet on the canal. Where on the canal should this common outlet be chosen so that the total length of pipe is as short as possible?

Let A and B be distant a, b respectively from the canal bank and distant c apart.

Model: We have a common outlet C and pipes CA, CB (which will clearly be straight lines) (see figure 5.1[*a*]).

Understanding: $AC+CB$ to be minimized.

Inductive investigation: Specialization gives:

(i) $a = b$ will need C symmetrically placed halfway, which leads to $l = (c^2+4a^2)^{\frac{1}{2}}$.

(ii) $b = 0$ which clearly gives C at $B \Rightarrow l = c$.

(iii) AB in line perpendicular to the canal, which gives $c = a-b$ and $l = a+b$.

Studying these three results may suggest the conjecture $l = (c^2+4ab)^{\frac{1}{2}}$. Unfortunately the conjecture does not suggest a method of approach unless we happened to rearrange it as $l = [c^2+(a+b)^2-(a-b)^2]^{\frac{1}{2}}$.

Analogy: The diagram should turn our thoughts to reflection in a mirror because we know that light always takes the path of least time in going from one point A to another

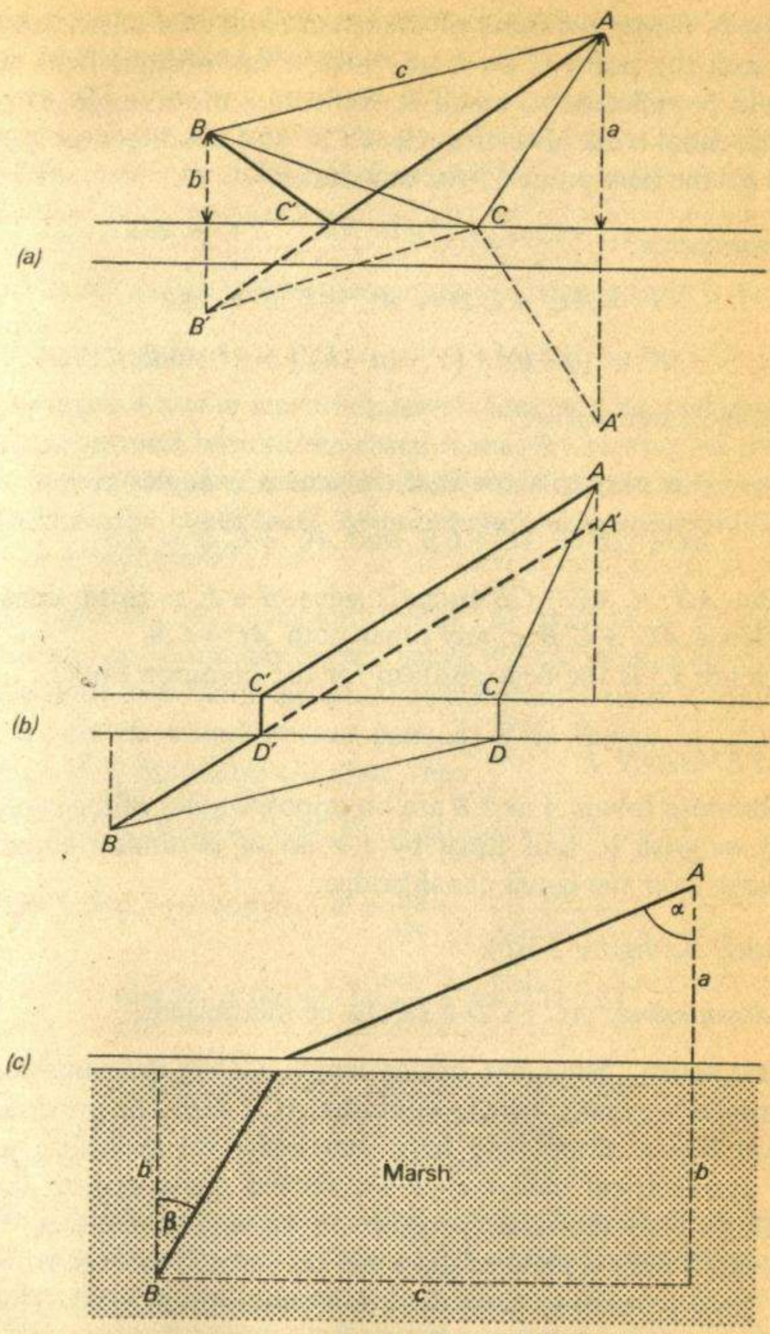


Fig. 5.1

point B. Hence we think of the canal bank as a mirror and we seek the point C on it at which a ray of light from A would be reflected to reach B. Reflecting the triangle ABC in the canal bank gives triangle $A'CB'$ and the direction AB' will fix the best point C' for our outlet.

Confirmation:

$$l' = AC'+C'B = AC'+C'B' = AB'$$

$$l'^2 = (a+b)^2+[c^2-(a-b)^2] = c^2+4ab$$

as conjectured earlier.

Proof: It is easy to show that congruent triangles give

$$AC+CB = AC+CB' \text{ and } AC'+C'B = AB'$$

But $AB' < AC+CB$ (since 2 sides of a $\Delta >$ third side)
Hence $AC'+C'B <$ any other path $AC+CB$.
Hence C' is the best position for the common outlet.

EXTENSION 1

Suppose towns A and B are on opposite sides of the canal and we wish to join them by a road of minimum length passing over the usual canal bridge.

Model: As figure 5.1(b).

Understanding: $AC+CD+DB$ to be minimized.

Investigation: Since the bridge length CD is the same for all routes we really have to minimize $AC+DB$. The previous result seems of no help here, but when we specialize to the symmetrical case $b = a$ we should again realize the analogy with the same problem in an optical context. If the canal were a plate of glass the rays would emerge from the glass parallel to their direction entering the glass. This suggests we need DB parallel to AC.

Hence we mark off AA' equal to the width of the canal

and join $A'B$ to cut the further bank at D'. Then D' is one end of the required bridge.

Proof:

$$AC'+D'B = A'D'+D'B' = A'B$$

$$AC+DB = A'D+DB \text{ which is greater than } A'B.$$

Hence $AC'+D'B$ will be the minimum distance from A to B.

EXTENSION 2

Villages A and B are on opposite sides of a narrow canal but the ground beyond the canal is marshy so that the cost per mile of making a road is £x from A to the canal and £kx from the canal to B. Where should the road be built to be least expensive?

Investigation: We probably think of this as a routine question on minimum values using the calculus. Let A, B be distant a, b from the canal and distant c apart parallel to the canal. Consider any path ACB as shown in figure 5.1(c). If C represents the cost, then

$$C = x(AC)+kx(CB) = x(a \sec \alpha+kb \sec \beta)$$

where $a \tan \alpha+b \tan \beta = c \Rightarrow \dfrac{d\beta}{d\alpha} = -\dfrac{a \sec^2 \alpha}{b \sec^2 \beta}$

$$dC = x\left(a \sec \alpha \tan \alpha - ka\frac{\sec^2 \alpha}{\sec^2 \beta} \sec \beta \tan \beta\right)$$

$$\frac{dC}{d\alpha} = 0 \text{ when } \sec \alpha \tan \alpha \sec^2 \beta = k \sec \beta \tan \sec^2 \alpha$$

hence $$\sin \alpha = k \sin \beta$$

This, together with relation $a \tan \alpha+b \tan \beta = c$ will enable us to determine the values of α and β, although the solution will be complicated.

The result $\sin \alpha = k \sin \beta$ should cause us to think of the problem in the optical context—as a ray of light from A to

B which is refracted towards the normal after crossing the canal. We simply think of 1 and k as the refractive indices on the different sides of the canal. Hence the usual geometrical construction given in elementary textbooks on optics will enable us to construct the required path $AC'B$.

Many other extensions are possible such as:

(*a*) Suppose a very hilly wedge-shaped piece of land lay between the two towns, necessitating a tunnel at great cost per mile. Again the optical context gives us a quick solution by thinking of the passage of light through a triangular prism;

(*b*) If we lived in Holland we would often find two or more parallel canals between the two towns. To find the shortest distance we simply generalize Extension 1.

PROBLEM 2. THE DERANGEMENT PROBLEM

In how many ways can the set of numbers $\{1, 2, 3, \ldots, n\}$ be deranged so that no number occupies its natural place.

This important theorem in 'combinatorics' has an obvious model and a clear objective. No definite method suggests itself and so we try an inductive investigation.

Investigation: Specialization for $\{1\}, \{1, 2\}, \{1, 2, 3\}$, etc., gives the values

$$\Delta_1 = 0, \Delta_2 = 1, \Delta_3 = 2, \Delta_4 = 9, \Delta_5 = 44,$$

where Δ_r is the number of derangements of exactly r integers. A pattern is not too obvious but one conjecture is

$$\Delta_{n+1} = n(\Delta_n + \Delta_{n-1}).$$

Clearly it would be very tedious to confirm this by considering the next case $n = 6$ (because it has 318 ways), so we try to prove it directly.

Proof: Consider the set $\{a_1, a_2, a_3, \ldots, a_{n+1}\}$ and let the

a_{n+1} position be replaced by one of the others, say $a_k \Rightarrow n$ ways of choosing a_k.

Either the a_k position is filled by $a_{n+1} \Rightarrow \Delta_{n-1}$ derangements, or the a_k position is not filled by $a_{n+1} \Rightarrow \Delta_{n-2}$ derangements.

Hence $$\Delta_{n+1} = n(\Delta_n + \Delta_{n-1}).$$

We have now a *subsidiary* problem—to solve this non-linear difference equation to find Δ_n.

Insight: We will find a solution hard to get unless we suddenly realize that the equation can be *restated* in a different form which is *more accessible.*

Rearrange $$\Delta_n = (n-1)(\Delta_{n-1} + \Delta_{n-2})$$

as $$\Delta_n - (n)\Delta_{n-1} = -[\Delta_{n-1} - (n-1)\Delta_{n-2}].$$

This is now easy to solve because it has the form $F(n) = -F(n-1)$ which gives us $F(n) = (-1)^{n-2}F(2)$.

Hence $\Delta_n - (n)\Delta_{n-1} = (-1)^n(\Delta_2 - 1\Delta_1) = (-1)^n$.

This can be rewritten as

$$\frac{\Delta_n}{n!} - \frac{\Delta_{n-1}}{(n-1)!} = \frac{(-1)^n}{n!}.$$

If we write this for $n = n, n-1, n-2, \ldots, 2$ and add, we get

$$\frac{\Delta_n}{n!} - \frac{\Delta_1}{1} = n!\left(\frac{1}{2!} - \frac{1}{3!} + \ldots + \frac{(-1)^n}{n!}\right),$$

hence $$\Delta_n = n!\left(\frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} - \ldots + \frac{(-1)^n}{n!}\right),$$

since $$\Delta_1 = 0.$$

REVIEW

This solution seems unnecessarily complicated and in our review we try to find a simpler and shorter method.

We may think up the idea that Δ_n is the *number* of non-zero terms in the expansion of the determinant

$$\Delta_n = \begin{vmatrix} 0 & 1 & 1 & . & . & . & . & . & . & 1 \\ 1 & 0 & 1 & . & . & . & . & . & . & 1 \\ 1 & 1 & 0 & . & . & . & . & . & . & 1 \\ . & . & . & & & & & & & . \\ . & . & . & & & & & & & . \\ . & . & . & & & & & & & . \\ . & . & . & & & & & & & . \\ 1 & 1 & 1 & . & . & . & . & . & . & 0 \end{vmatrix}$$

where every diagonal principal element is zero and every other element unity. If we take $+$ as the sign of every term in the expansion, Δ_n will then be the value of this altered determinant, called a *permanent*. However, unless the properties of such permanents are known, this method again leads to the difference equation $\Delta_n = (n-1)(\Delta_{n-1}+\Delta_{n-2})$.

We may, however, suddenly realize that the total number of derangements on n elements is the sum of the number of ways with exactly r misplaced, as r varies from 1 to n. This gives us a new equation:

$$n! = 1+\binom{n}{1}\Delta_1+\binom{n}{2}\Delta_2+\ldots+\binom{n}{n}\Delta_n.$$

This does not look very promising unless we suddenly get a spot of inspiration—doesn't the right-hand side remind us of the binomial theorem? This *analogy* leads us to an easy solution when we see that the mapping $\Delta_r \to \Delta^r$ gives an isomorphism between the above right-hand side and the expansion of $(1+\Delta)^n$.

Hence $r! \to (1+\Delta)^r$ where Δ^k represents Δ_k.

$$\Delta_n \to \Delta^n = [(1+\Delta)-1]^n = \sum_{r=0}^{n}(-1)^r\binom{n}{r}(1+\Delta)^{n-r}$$

$$\to \sum_{r=0}^{n}(-1)^r\binom{n}{r}(n-r)!$$

$$= n!-\binom{n}{1}(n-1)!+\binom{n}{2}(n-2)!-\ldots)$$

$$\text{hence } \Delta_n = n!\left(1-\frac{1}{1!}+\frac{1}{2!}-\frac{1}{3!}+\ldots+\frac{(-1)^n}{n!}\right).$$

EXTENSION 1: What is the probability of a complete derangement? Since the total number of ways of arranging n objects in a row is $n!$, the probability is clearly given by

$$P(n) = 1-\frac{1}{1!}+\frac{1}{2!}+\frac{1}{3!}+\ldots+\frac{(-1)^n}{n!}.$$

We surely notice that the right-hand side is the first n terms in the expansion of e^{-1}. Hence as n gets large, we get the very surprising result that

$$P(n) \to 1/e.$$

EXTENSION 2: If n letters are written and n envelopes correctly addressed, in how many ways would exactly four of the letters be misposted? This is simply one of the several interesting ways of *restating* the problem. We can choose 4 letters from 8 in $\binom{8}{4}$ ways and can then mispost these 4 letters in Δ_4 ways.

$$\Delta_4 = 4!\left(\frac{1}{2}-\frac{1}{6}+\frac{1}{24}\right) = 12-4+1 = 9.$$

Hence the number of ways is $70\times 9 = 630$.

One honours student of mine gave an intriguing solution to the original problem as follows:

$$\Delta_n = n!-\left[\!\!\left[\binom{n}{1}(n-1)!-\left\{\binom{n}{2}(n-2)!-\left[\binom{n}{3}(n-3)!\ldots\right.\right.\right.\right.$$

$$\left.\left.\left.\left.-\left[\binom{n}{n-1}1!-\binom{n}{n}0!\right]\right\}\ldots\right]\!\!\right].$$

When asked to amplify this solution he admitted it was a swindle. Actually, careful thought shows that there is more to this solution than he realized, and there is a method of solution that obtains the answer in this form.

PROBLEM 3. CHANGING MONEY

In how many ways can change be given for one dollar in the current coinage of 1, 5, 10, 25, 50 cents called a 'cent', 'nickel', 'dime', quarter', and 'half' respectively.

Writers who discuss money-changing problems usually use the fact that the number of ways of changing a sum valued n cents is the coefficient of x^n in the expansion of

$$(1+x+x^2+\ldots)(1+x^5+x^{10}+\ldots)(1+x^{10}+x^{20}+\ldots)$$
$$(1+x^{25}+x^{50}+\ldots)(1+x^{50}+x^{100}+\ldots).$$

This is changed to the equivalent expression:

$$1/(1-x)(1-x^5)(1-x^{10})(1-x^{25})(1-x^{50}),$$

and the resulting algebra is particularly horrible! We suggest that the following method is simpler and shorter and much more accessible for larger sums than one dollar.

Let us write $N_r(x)$ for the number of ways of changing a sum valued x cents with only the first r coins of the set 1, 5, 10, 25, and 50. Clearly any odd cents less than 5 will not affect the number of ways of receiving change and so it suffices to consider sums values $5k$ cents, where k is an integer.

$$\left[\frac{x}{5}\right] = k \text{ has its usual meaning.}†$$

LEMMA 1: $N_1(x) = 1$ and $N_2(x) = 1+\left[\frac{x}{5}\right]$ are obvious.

† $[x]$ is the greatest integer not greater than x.

LEMMA 2: $N_3(x) = \left(1+\left[\frac{x}{10}\right]\right)^2$ if k is even,

$$= \left(1+\left[\frac{x}{10}\right]\right)\left(2+\left[\frac{x}{10}\right]\right) \text{ if } k \text{ is odd.}$$

Proof: We can use 0 nickels, 1 nickel, 2 nickels, and so on:

$$k \text{ even} \Rightarrow N_3(5k) = 1+N_2(10)+N_2(20)+\ldots+N_2(5k)$$
$$= 1+3+5+\ldots+(k+1) \text{ by Lemma 1}$$
$$= (1+k/2)^2 = \left(1+\left[\frac{x}{10}\right]\right)^2.$$

k odd $\Rightarrow$ as above we have

$$N_3(5k) = N_2(5)+N_2(15)+N_2(25)\ldots+N_2(5k)$$
$$= 2+4+6+\ldots+(k+1)$$
$$= \tfrac{1}{4}(k+1)(k+3)$$

and $\quad k = 1+2\left[\frac{x}{10}\right]$ when k is odd

$$N_3(5k) = \left(1+\left[\frac{x}{10}\right]\right)\left(2+\left[\frac{x}{10}\right]\right) \text{ when } k \text{ is odd.}$$

This can be written as $N_3(5k) = \left(1+\left[\frac{x}{10}\right]\right)^2+\left(1+\left[\frac{x}{10}\right]\right)$ for odd k.

For example, $N_3(50) = (1+5)^2 = 36$; $N_3(75) = 8\cdot 9$ or $8^2+8 = 72$.

We now return to the original problem.

Solution: One dollar = 100 cents.

$N_5(100) = N_4(100)+N_4(50)+1$, since we can use 0, 1, or 2 half-dollars.

$$= 1+[1+N_3(25)+N_3(50)]+$$
$$[1+N_3(25)+N_3(50)+N_3(75)+N_3(100)]$$
$$= 3+2[N_3(25)+N_3(50)]+[N_3(75)+N_3(100)]$$

$$= 3+2(3 \,.\, 4+6^2)+(8 \,.\, 9+11^2) = 3+2(48)+193$$
$$= 292.$$

EXTENSION 1: An explicit formula for $N_4(25k)$.
$N_4(25k)$ will again have distinct values for k even and k odd.

k even:

$$N_4(25k) = 1+N_3(25)+N_3(50)+N_3(75)+\ldots+N_3(25k)$$
$$= 1+(3 \,.\, 4)+6^2+(8 \,.\, 9)+11^2+(13 \,.\, 14)+\ldots +(5k+1)^2$$
$$= 1+\sum_1^k (5r+1)^2+\sum_1^k (5r-2)(5r-1)$$
$$= 1+\frac{k}{24}(50k^2+135k+106) \text{ on evaluation.}$$

Similarly when k is odd, the same method gives:

$$N_4(25k) = (k+1)(50k^2+85k+21)/24.$$

EXTENSION 2: In how many ways can we change £1 in the decimal currency (a) using the 5, 10, 50 coins and (b) using the 1, 2, 5, 10, 50 coins.

The easy part (a) is inserted to suggest a solution by the 'horrible' algebraic method quoted above, but we should at least simplify by putting $y = x^5$ in the algebraic expressions. Part (b) can be solved by the method used above: start with lemmas $N_1(x) = 1$, $N_2(x) = [x/2]+1$, and then evaluate $N_3(5k)$ for k even and k odd.

PROBLEM 4. THE CORRIDOR PROBLEM

What is the length of the longest pole that can be carried horizontally round a right-angled corner where two straight corridors of width a and b meet? (See figure 5.2(a).)

Clearly when carried round the corner, the longest pole will be the one that touches the inner corner C and just scrapes the walls at the ends A and B.

Model: $l = a \sec\theta+b \operatorname{cosec}\theta$.

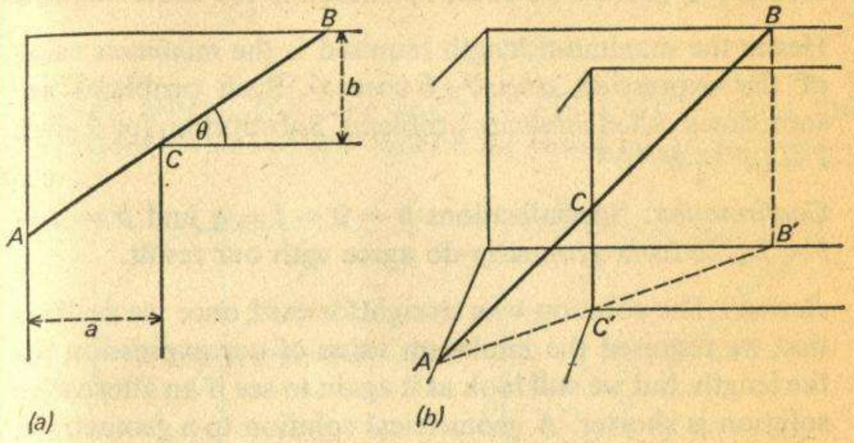


Fig. 5.2

Understanding: Maximum l required.

Solution: $\dfrac{dl}{d\theta} = a \sec\theta \tan\theta - b \operatorname{cosec}\theta \cot\theta$.

$$\frac{dl}{d\theta} = 0 \text{ when } a \sec\theta \tan\theta = b \operatorname{cosec}\theta \cot\theta$$
$$\Rightarrow \tan^3\theta = b/a.$$

Hence $\tan\theta = b^{\frac{1}{3}}/a^{\frac{1}{3}}$ leads to stationary value for l.

$$\frac{d^2 l}{d^2\theta^2} = a(\sec^3\theta+\sec\theta\tan^2\theta)+ b(\operatorname{cosec}^3\theta+\operatorname{cosec}\theta\cot^2\theta) > 0 \text{ for all acute } \theta.$$

Hence l given by this value of θ is a *minimum*.

Thus we seem to have a paradox—the largest length on geometrical intuition is now a minimum length. The explanation of this paradox is that we have drawn one par-

ticular value of θ and made this the particular pole for that angle. What we should have written is:

possible rod length $l \leqslant a \sec\theta + b \operatorname{cosec}\theta$ for all acute angles θ

Hence $l \leqslant (a \sec\theta + b \operatorname{cosec}\theta)$ minimum for acute angles θ

Hence the maximum length required is the *minimum* value of the expression $a \sec\theta + b \operatorname{cosec}\theta$. Such problems are sometimes called *minimax* problems. Substitution for θ gives $l = (a^{2/3}+b^{2/3})^{3/2}$.

Confirmation: Specializations $b = 0 \Rightarrow l = a$ and $b = a \Rightarrow l = 2\sqrt{2}a$ from symmetry do agree with our result.

Review: The solution was straightforward once we realized that we required the minimum value of our expression for the length, but we still look at it again to see if an alternative solution is shorter. A geometrical solution to a geometrical problem seems more natural and this is how an applied mathematician would solve it.

Solution: Any pole above a certain length will jam as it passes round the corner touching the walls at A, B, and C. The longest pole to pass right round must be moving tangentially to the walls at A, B, and C. Hence the perpendiculars to the walls at A and B will give the instantaneous centre K and so KC must be perpendicular to the pole AB. Simple geometry of similar triangles now leads to the above solution.

EXTENSION 1: What is the longest pole that can be carried round the right-angle corner if the walls have height h?

Investigation: An examination of figure 5.2(b) shows that the longest pole AB (whose projection on floor makes angle θ with the horizontal wall) at a given value θ will be the pole that touches the two outer walls at A and B

and the inner wall at C. Our *longest* pole carried round the corner will be the *minimum* value of that AB.

Solution: The right-angled triangle $BB'A$ requires that the projection of AB on the floor should touch the inner wall at C'.

$$l^2 = h^2 + (AB')^2.$$

Hence longest pole has length $\{h^2+(a^{2/3}+b^{2/3})^{3/2}\}^{1/2}$ since the maximum value of AB' will be the result in our problem.

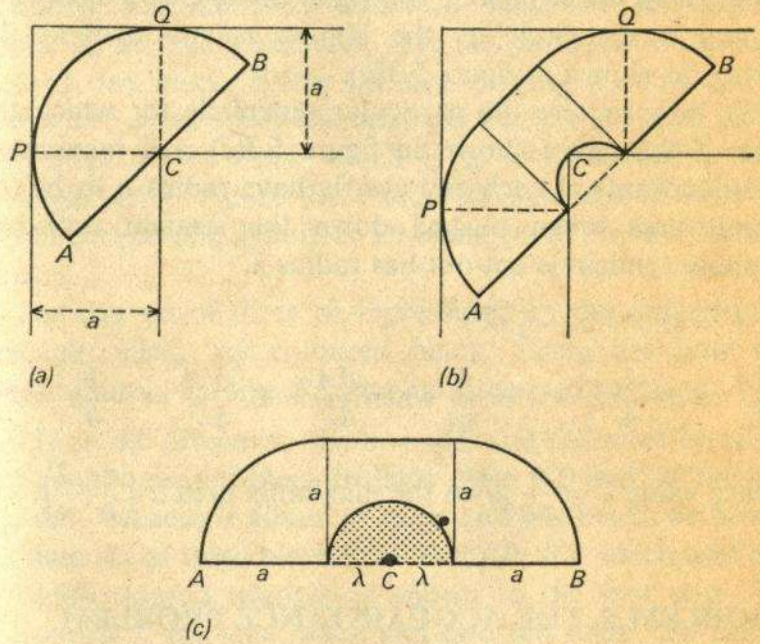


Fig. 5.3

EXTENSION 2: What is the largest flat metal sheet we can push on the floor around the right-angled corner when the corridors have equal widths.

Investigation: The largest sheet must clearly be in contact with the outer walls and with the inner wall for acute angles θ once it starts to turn around the corner. Thus in

figure 5.3(a) we need $CP = CQ = a$ for all acute angles θ. This seems only possible if the lemma is semicircular.

Hence largest area is $\frac{1}{2}\pi a^2$.

Although the argument seems sound, it is not correct. What we have not used is the property that, although the largest sheet must pivot around the inner wall at C, the point of contact of the base of the sheet need not be the *same* point C for all θ. A little insight is now needed to realize that the right-angle at C gives us a vital clue. Since the angle in the semicircle is a right-angle, C could traverse around a semicircle as the lamina rotates around the corner, as shown in figure 5.3(b).

We now require the particular semicircle for which the area of the figure shown in figure 5.3(c) is a maximum. The quadrants at each end clearly have radius a to be the largest area when pushed down the straight corridor. Suppose semicircle cut out has radius λ.

$$A = \tfrac{1}{2}\pi a^2 + a\lambda - \tfrac{1}{8}\pi\lambda^2.$$

$$\frac{dA}{d\lambda} = a - \frac{1}{4}\pi\lambda \Rightarrow \lambda = \frac{4}{\pi a} \text{ when } \frac{dA}{d\lambda} = 0. \quad \frac{d^2A}{d\lambda^2} = -\frac{1}{4}\pi < 0.$$

Hence value $\lambda = \dfrac{4}{\pi a}$ gives the maximum area $a^2\left(\dfrac{\pi}{2} + \dfrac{2}{\pi}\right)$.

PROBLEM 5. THE ACQUAINTANCE PROBLEM

If six people meet at a party, show that there must be at least two sets of three who are either complete strangers or mutual acquaintances.

Investigation: We soon find that the method of exhaustion is not feasible as we find difficulty in cataloguing the various possibilities. We clearly require a representation of a reciprocal property that applies to a group of six, two at a time. A little thought should suggest a geometrical model—a

hexagon where the six vertices represent the six people or a square with two symmetrical points selected inside the square. The relation between them is represented by colouring the line joining them—coloured black if they are strangers and red if they are mutual acquaintances.

Model: The six vertices of a hexagon and two colours.

Understanding: We have to show that there always exists at least two self-coloured triangles (all sides of same colour) with vertices among the six vertices of the hexagon.

Solution: Consider any vertex A. Since five lines radiate from A, at least three AX, AY, AZ must have the same colour, say black. Either the lines joining X, Y, Z are all red which gives a self-coloured red triangle XYZ or one of the sides XY is black which gives a self-coloured black triangle AXY. Hence there must always exist at least *one* set of people who are either all strangers or else all acquaintances.

Let this set of three be represented by the self-coloured triangle ABC, say coloured black. There are now two possibilities—the line BE may be either red or black.

(a) Line BE drawn in black $\Rightarrow$ AE and CE must be red to avoid a second black triangle. Also BD and BF must be red, because if either of them (BX) is black we have a free set of three black lines BA, BE, BX which lead to a self-coloured triangle as shown in the first part. The vertex E has one black and two red lines. We cannot draw either ED or EF red as this again gives a free set of three red lines from E leading to a self-coloured triangle. Hence ED and EF must be black. Then, as shown in figure 5.4(a), we cannot join DF without making a new self-coloured triangle.

(b) Line BE is drawn in red. The lines BD and BF cannot both be the same colour or that would give a free set of three red lines BD, BE, BF or three black lines BA, BF,

BD radiating from *B* which again lead to a self-coloured triangle. Hence one is red and one is black and the symmetry shows the choice is immaterial—take *BD* red and *BF* black. Lines *AF* and *CF* must be red to avoid a new self-coloured black triangle. Vertex *F* now has two

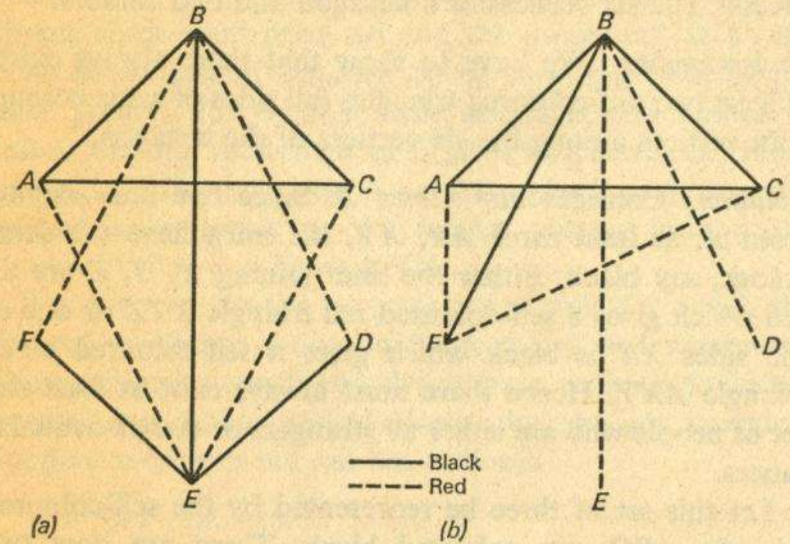


Fig. 5.4

red and one black line radiating from it and the same argument as in (*a*) shows we must get a new self-coloured triangle because of a free set of three same-coloured lines from *F*—see figure 5.4(*b*).

In both cases at least two self-coloured triangles must exist and so there are at least two sets of three people who are either strangers or mutual acquaintances.

EXTENSION 1: What conditions lead to *exactly* two self-coloured triangles?

There are two possibilities—the two sets may have the same colour or they may have opposite colours. Interested readers should find it quite easy to show that figures are possible.

(i) With *exactly* two triangles of the same colour in all possible cases (i.e. with 0, 1, or 2 vertices in common),

(ii) With *exactly* two triangles of opposite colours if and only if the two triangles have exactly one vertex in common.

EXTENSION 2: Can we generalize?

We soon see that five points need have no self-coloured triangle and, with a great deal of effort, that seven points always have at least four self-coloured triangles. The general result was first proved by A. W. Goodman in *A.M.M.*, no. 66 (1959) in the form: a party of n people must contain at least k sets of three, all strangers or all mutual acquaintances, where k is given by

$$k = \binom{n}{3} - \left[\frac{n}{2}\left(\frac{n-1}{2}\right)^2\right],$$

the square bracket being the usual integer function.

EXTENSION 3: THE GAME OF SIM

This model hexagon can be used for an interesting game called 'SIM' after its inventor G. Simmons. Six dots are marked on a sheet of paper and two players take turns to join one pair of dots with a straight line, one player using black and one red. The first player forced to complete a triangle of his own colour loses. Analysis has shown that the second player can always win if he chooses the right strategy, but this strategy is highly complicated and among skilful players wins and losses are equally likely.

PROBLEM 6. THE TWELVE COINS

Given twelve coins, eleven genuine and one counterfeit, find in three weighings on an equal-arms balance (without using any weights) which is the counterfeit coin and whether it is heavier or lighter than a genuine one.

Investigation: Many solutions, some quite complicated, have been provided for this famous war-time problem, but the following method seems very simple and direct. We need a representation which includes at least twelve distinct elements and such that the conjugate (or exact opposite) of any element is not included among the twelve chosen—this is to allow for the heavier or lighter requirement. One obvious representation is a vector with three components (L, R, O) where L, R, O stand for 'left', 'right', and 'omitted' respectively. This gives $3^3 = 27$ elements, which reduce to 13 distinct elements when we pair off each member with its conjugate. The extra one is the (O, O, O,) representation which is discarded in view of the heavier or lighter requirement. The following instructions could be given to a child, and all he need say is the answer—which is the heavier arm at each of the weighings.

Solution

1. Number the coins 1–12.
2. Place the coins on the balance in following order:

	Right		*Left*
First weighing:	1, 2, 3, 4	*v.*	5, 6, 7, 8
Second weighing:	3, 4, 5, 6	*v.*	8, 9, 10, 11
Third weighing:	1, 5, 7, 9	*v.*	4, 8, 10, 12

3. Now read off the counterfeit coin from the following table:

1	2	3	4	5	6	7	8	9	10	11	12
R	R	R	R	L	L	L	L	—	—	—	—
—	—	R	R	R	R	—	L	L	L	L	—
R	—	—	L	R	—	R	L	R	L	—	L

If the result fits a column in the table, the coin at the top of that column is the counterfeit coin and the coin is too

heavy; if the conjugate result occurs, the coin is too light. For example, (R,R,L) identifies coin 4 as too heavy; (L,L,R) identifies coin 4 as too light.

Review: Is there a simpler solution? We do not think that a child could manage the continuous reasoning needed during a direct solution, but interested readers can work out a schedule which depends at each stage on the information obtained from the previous weighing. For example, if the first weighing *balanced*, the dud coin must be one of the set 9, 10, 11, and 12. Now we can identify the dud by continuing:

Second weighing		*Third weighing*
	if balance,	1 *v.* 12
1, 9 *v.* 10, 11		
	if not balance,	1, 2 *v.* 9, 10

EXTENSION 1: Can we generalize?

What is the maximum number of coins from which to determine a dud coin and whether too light in exactly n weighings? The above method will extend to the general case, but our vector now has n components. This will give $\frac{1}{2}(3^n-1)$ distinct elements each of which can be associated with a chosen coin. The first weighing most separate the coins into three equal groups and so we subtract one coin to get the maximum number $\frac{1}{2}(3^n-3)$. For subsequent weighings the extra coin can be accepted because we then have some extra information—a genuine coin (see Extension 3).

EXTENSION 2: AN OVERWEIGHT DUD

If we are told that the dud coin is overweight, then we can determine the dud coin from a set of $3^3 = 27$ coins in three weighings. The weighing of 9 coins against 9 other coins determines the set of 9 which includes the dud coin. The

second weighing of 3 against 3 will determine the set of 3 coins which include the dud. The last weighing of 1 against 1 identifies the dud.

In the general case of n weighings, we can start with 3^n coins and our first weighing is then 3^{n-1} coins against another set of 3^{n-1} coins, etc.

EXTENSION 3: Suppose we are also given a *genuine* coin as well as the set of 12 coins.

We find that the original problem can now be solved with a set of 13 coins from which to determine the dud coin and its type. In the general case the provision of an extra coin which is guaranteed genuine enables this problem to be solved with $(n+1)$ coins instead of n coins.

These problems were selected to illustrate important points either in the inductive investigation or in the over-all strategy as well as to show how interesting extensions can be formulated. Problem 1 shows how the use of *analogy* extended from the problem to the various extensions. Problem 2 illustrates different methods of approach to a problem. Problem 3 shows that the simple direct method of exhaustion may still be the shortest way of tackling some problems. Problem 4 illustrates the distinct types of extensions, an obvious practical one and an interesting theoretical one. Problem 5 provides an example where half the battle is the finding of a good model. Problem 6 had to be included as the most popular problem of the last thirty years!

Before ending this chapter on the '*review*' of solutions we should review the problems solved in the earlier chapters. Immediately alternative solutions spring to mind and interesting extensions suggest themselves. We will consider just two of them.

A. The Four Knights Problem in Chapter 2, Example 2

The 'curious man' mentioned in Chapter 1 tried this problem and gave me a remarkably neat and simple solution. If we number the outside squares from 1 to 8, we have white knights x_1 and x_3 on squares 1 and 3, and black knights y_5 and y_7 on squares 5 and 7.

Solution: Perform the mapping $S(x_r) \to x_{r+3}$ (mod 8) four times!

$$\begin{aligned} S^4(x_1,x_3,y_5,y_7) &= S^3(x_4,x_6,y_8,y_2) = S^2(x_7,x_1,y_3,y_5) \\ &= S(x_2,x_4,y_6,y_8) = x_5,x_7,y_1,y_3. \end{aligned}$$

Readers may find it difficult to see *why* it works, but it is really a combination of the knight's move and four rotations through 90°.

B. The Derangement Problem of this chapter

When looking over this problem a much simpler expression for the final result became apparent. The number is $[n!/e]$ when n is odd and $1+[n!/e]$ when n is even.

Proof: $\Delta_n = n!\left(\frac{1}{2!}-\frac{1}{3!}+\frac{1}{4!}-\dots+\frac{(-1)^n}{n!}\right) = \frac{n!}{e}-R_n,$

where remainder

$$R_n = n!\left[\frac{(-1)^{n+1}}{(n+1)!}\dots\right] \text{ so that } |R_n| < \frac{n!}{(n+1)!} = \frac{1}{n+1} < 1.$$

A study of the signs shows that Δ_n is less than $\frac{n!}{e}$ when n is odd and greater when n is even. Since the difference is less than one, we have

$$\Delta_n = \left[\frac{n}{e}\right] \text{ when } n \text{ odd, } 1+\left[\frac{n}{e}\right] \text{ when } n \text{ even.}$$

In concluding we hope that every reader has found something of special interest in the preceding pages and

that occasionally a suggestion made has prompted serious reflection on the art of problem-solving. We trust that readers will attempt some of the problems in the following collection and that, when the problem is not a routine or a manipulative example, they will tackle it in the spirit in which the book is written.

Collection of 100 Exercises

The following collection of exercises is not arranged in order of difficulty and does not contain many routine problems of the standard types to be found in the usual school textbooks on the relevant topics. Most of them are well within the ability and range of knowledge of the average student who has covered the normal A-level course in mathematics. Some need only very simple thought and have short solutions. A few are designed not so much to develop routine techniques as to stimulate inventive ability and critical ability.

It is sometimes more instructive in manipulative examples to solve a problem by two different methods rather than to work out several examples of a similar type by one method, although the latter is necessary when learning new techniques. Some of the exercises (such as numbers 6, 16, 26) are inserted specifically for this purpose.

(Remember that the starred questions are more difficult either because of the knowledge required or because of the depth of insight needed.)

1. Prove that every right-angled triangle with each side an exact integer must have at least one side a multiple of 5.

2. For what real values of k do the expressions

$$x^2+kx+2 \text{ and } 2x^3+kx-4$$

possess a common factor.

3. Show that $x^x = 2$ possesses exactly one real root and find its value to two decimal places.

4. Find integers λ and μ such that

$$4{,}189\lambda + 2{,}183\mu = 59.$$

Are these integers unique?

5. A convex polygon has n sides. Find the number of diagonals and the maximum number of distinct points of intersection in the plane of these diagonals where we exclude the vertices of the polygon.

6. Prove that the sum of the first n natural numbers

$$1+2+3+\ldots+n = \tfrac{1}{2}n(n+1),$$

(i) by direct deduction, (ii) by mathematical induction, (iii) by the method of contradiction, and (iv) by a diagrammatic proof.

7. Given $3^3+4^3+5^3 = 6^3$, find a *simple* method of generating new sets of integers (other than direct multiples of those given) to satisfy the relation $a^3+b^3+c^3 = d^3$.

8. Find rational factors of

(i) x^4+2x^2+9.

(ii) $x(x+1)(x+2)(x+3)-24$.

9. Prove that the square of every odd integer >1 is of form $8n+1$ and hence deduce 17^n-1 is divisible by 8 for all positive integers n. Prove also by one other method.

10. If we assume the result

$$\sum_1^\infty \frac{1}{r^2} = \frac{1}{6}\pi^2,$$

find the sum to infinity of the following convergent series

(i) $1+\frac{1}{3^2}+\frac{1}{5^2}+\frac{1}{7^2}+\ldots$

(ii) $1-\frac{1}{2^2}+\frac{1}{3^2}-\frac{1}{4^2}+\ldots$

11. Evaluate $\sum_{j=1}^{n}\sum_{i=1}^{j}\sum_{r=1}^{i} 1$.

12. A number of six digits is divisible by 37. Prove that the sum of the two 3-digit numbers forming it must be divisible by 37. Find also a similar rule for 8-digit numbers divisible by 37.

13. P is a given point inside an acute angle AOB. Construct a straight line through P to cut off the triangle of minimum area from the arms of the angle.

14. If the integers $1,2,3,\ldots,n$, are arranged in any order $a_1,a_2,\ldots,a_n$, show that the product

$$(a_1-1)(a_2-2)(a_3-3)\ldots(a_n-n)$$

must be even (or zero).

15. Show that the radian measure of any acute angle α satisfies $\alpha < \frac{1}{2}(\sin\alpha+\tan\alpha)$.

16. If

$$\frac{a-b}{1+ab} = \frac{c-d}{1+cd},$$

prove both by algebra and by trigonometry that

$$\frac{a-c}{1+ac} = \frac{b-d}{1+bd}.$$

Which of these methods gives the easier solution of the following new problem?

If $x+y+z = xyz$, prove that $\sum x(1-y^2)(1-z^2) = 4xyz$.

17. Explain why each of the following multiplications could be done mentally at sight:

(i) $143 \times 927 = 132{,}561$.

(ii) $1{,}667 \times 681 = 1{,}135{,}227$.

Try in each case to find a rule for multiplying any 3-digit number *abc* by 143 or by 1,667.

18. Show that the quadratic expression ax^2+bx+c takes on integral values for every integer x if and only if $2a$, $a+b$, and c are each integers.

19.* A triangle ABC has $AB = AC$ and perimeter $= 5BC$. A point P is chosen arbitrarily inside the triangle and Q is the mid-point of AP. Prove that the sum of the lengths of the perpendiculars from P to the sides of the triangle equals the length of the perpendicular from Q to BC.

20. The inscribed circle of triangle ABC has radius 4 cm and touches the side BC at P where $BP = 6$ cm, $CP = 8$ cm. Find the length of the shortest side of the triangle.

21. Show that the determinant

$$\begin{vmatrix} a & c & b \\ b & a & c \\ c & b & a \end{vmatrix} = a^3+b^3+c^3-3abc.$$

Hence deduce that $(a^3+b^3+c^3-3abc)(x^3+y^3+z^3-3xyz)$ can be put in the form $A^3+B^3+C^3-3ABC$.

22.* A circle has chord AB with mid-point O. Any two chords POQ and ROS are drawn with P and R on same side of AB. The lines PS and QR cut AB at H and K. Prove that O is also the mid-point of HK.

23. A ladder of length 40 ft with its base on level ground and its top against a vertical wall rests at right angles to and in contact with a cubical box of side 10 ft. If the box rests against the wall, how high up the wall does the ladder reach?

24. Given that $A_n = \sum_{r=0}^{n} a_r$, $b_n = \sum_{r=0}^{n} a_r a_{n-r}$, $B_n = \sum_{r=0}^{n} b_r$,

prove that $$B_n = \sum_{r=0}^{n} a_r A_{n-r}.$$

25.* Solve the difference equation

$$I_{n+2}-4I_{n+1}+4I_n = n.$$

26. Show by at least two methods that, given that a triangle ABC has $\hat{A} = 2\hat{B}$, then $a^2 = b(b+c)$.

27. Assuming Euler's result for convex sides of three dimensions, that $E+2 = F+V$ (where V, E, F are the number of vertices, edges, and faces of the solid), find a similar result for a four-dimensional solid. What is the equivalent result in n dimensions?

28. If M is the lowest common multiple and D the highest common factor of three positive integers, prove that the product MD cannot exceed the product of the three integers.

29. If n is a positive integer, find the number of positive integer solutions (excluding zero solutions) of the equation: $2x+y+z = 2n$.

30. Two altitudes of a triangle have lengths 3 cm and 5 cm. Show that the third altitude lies between 7·5 cm and 1·875 cm.

31. A circle is intersected by n distinct straight lines, none of which is a tangent nor do any pair of lines intersect on the circumference. Find:

(i) The least number of regions into which the circle can be divided.

(ii) The least number of regions if every pair of lines must intersect.

(iii) The greatest number of regions if no two lines are parallel and no three lines are concurrent.

32. Evaluate 3^{3^3} and show that 7^{7^7} is a number with more than 695,000 digits.

33. Decide, without using any tables, which is the larger number of 2^{85} and 3^{53}.

34. A gardener wished to plant a number of plants equi-spaced in rows of equal length, using more than one row. He found that the least number of rows was seven, since with a lesser number of rows the number of plants left over was always one less than the number of rows. What is the smallest number of plants for which this is possible?

35. A theorem on limits proves that, if $\{u_n\}$ is a sequence converging to a non-zero limit l, then

$$\frac{1}{n}\sum_{r=1}^{n} u_r$$

also converges to the same limit l. See how useful this result is by testing the case $u_n = 1/n$.

Take $u_n = n \log \{n/(n+1)\}$ and find values of l and $\sum u_r$. Then use the above theorem to deduce that

$$\frac{(n!)^{1/n}}{n+1} \to \frac{1}{e} \text{ as } n \to \infty.$$

Show that this result is compatible with Stirling's approximation for large $n!$:

$$n! \sim (2\pi n)^{\frac{1}{2}}\left(\frac{n}{e}\right)^n.$$

36. Let $f: A \to B$ be a mapping from A to B. Prove that f is invertible if and only if f is bijective, i.e. *one–one* and *onto*. [*Note:* f is invertible means the existence of a mapping $g: B \to A$ such that $gf = I_A$ and $fg = I_B$.]

37. A room is 30 ft long, 12 ft wide, and 12 ft high. A spider squats in the middle of one end wall 1 ft from the floor and a fly perches in the middle of the opposite end wall 1 ft from the ceiling. Find the shortest path the spider must crawl to get his breakfast.

38. An explorer camps at the base of a conical hill 1,800 ft high and 750 ft radius. He wishes to walk right round the hill, starting and ending at the camp. What is the length of the shortest route he can take?

39. If x^2+px+a divides exactly into x^4+qx^2+p, find the values of p and q in terms of a.

40. If I_n is a function of positive integer n, find I_n in each of the following cases, in each of which $I_1 = 1$.

(i) $I_n = nI_{n-1}$.

(ii) $I_n - I_{n-1} = n$.

(iii) $I_n + I_{n-1} = n$.

41. For what values of θ, if any, are the following satisfied:

(i) $2 \cos \theta + 3 \cos 2\theta = 6$.

(ii) $(1-\cos 2\theta)^2 + (1+\cos \theta)^2 = 0$.

(iii) $\cos 2\pi\theta = \sin \pi\theta$.

(iv) $\cot 2\theta + \tan \theta = \operatorname{cosec} 2\theta$.

42. Show that the line $(k-1)x+(k+1)y-2k = 0$ always passes through a fixed point for all values of k, and find the minimum length intercepted on this line by the circle $x^2+y^2 = 4$, as k varies.

43. Four spheres each of radius R lie on a horizontal table all touching one another. A fifth sphere is to be inserted to touch all these four spheres. Prove its radius is $R/(2+\sqrt{6})$.

44. 1 January 1956 was a Sunday. Prove that the first day of any century year cannot be a Sunday and find what day of the week will be 1 January 2100.

45. A farmer went to market in 1929 and bought exactly 100 animals at total cost of £100. If hens were 1*s*. each, pigs £4, and cows £20, how many of each did he buy?

46. A function f satisfies the relation

$$\mathrm{f}(x)+\mathrm{f}(y) = \mathrm{f}(x+y)+xy \text{ for all } x, y \in \mathbf{Z}.$$

If $\mathrm{f}(1) = 1$, find the value of $\mathrm{f}(n)$.

47. A convex lens is formed by the intersection of two spheres of radii 6 cm and 7 cm whose centres are 10 cm apart. Find the surface area of the lens. Why would the solution be simplified if the second sphere had radius 8 cm instead of 7 cm?

48. Show that at any party the number of guests who shook hands an odd number of times is even.

49. The circle $x^2+y^2 = 1$ has parametric coordinates $x = (1-t^2)/(1+t^2)$, $y = 2t/(1+t^2)$. Where the diameter $y = 0$ intersects the circle, we have $t = 0$ which implies $x = 1$. Hence this diameter cuts the circle only in the *one* point (1, 0). Explain this paradox.

50. Gauss proved that every prime number of the form $4k+1$ can be represented as the sum of two squares x^2+y^2. He gave a constructive existence type of proof to obtain

$$x = \frac{(2k)!}{2(k!)^2} \pmod{p} \text{ and } y = (2k!)x \pmod{p}.$$

Verify this result for $p = 17$ and 29. Is the result a practical one?

51.* Use the inductive method to obtain a conjecture for the position of points P and Q on the arms of a given angle AOB such that the curve joining PQ has minimum length and the area OPQ bounded by the arms and this curve has a given value.

52. A person on holiday noticed that it rained on 9 days and there were 10 clear mornings and 9 clear afternoons. When it rained in the morning it was clear in the afternoon. How long was his holiday?

53. If a and b are positive constants not equal to 1, solve:

(i) $(\log_x 2x)(\log_{10} x) = 10.$

(ii) $(\log_a x)(\log_b x) = \log_a b.$

54. A triangle of given area Δ has one angle equal to α. Find the lengths of the sides, if the side opposite the angle α is to be as short as possible.

55.* Prove that, among the positive numbers $a, 2a, 3a, \ldots, (n-1)a$ there is one that differs from an integer by at most $1/n$.

56. Here are two famous conjectures about some integer $n > 1$:

(i) There are only three squares of form $n!+1$.

(ii) There is a unique solution to $n!(n+1)! = k!$

Find the particular numbers referred to.

57. A circular piece of metal is cut out of a square piece of side a, and then a square piece of maximum size is cut out of the circular piece. How many circular pieces have been cut out when the final piece is a square of side $\frac{1}{8}a$.

58. A two-gallon radiator is filled with water. Four quarts are removed and replaced with pure anti-freeze liquid. After mixing, the process is repeated a second, third, and fourth time. How much water remains in the radiator after the fourth operation?

59. An integer a is such that the ten's digit of a^2 is odd. What is the unit digit of a^2?

60.* Given any two relatively prime integers a and b, show that the *largest* number not expressible in the form $ax+by$, where x and y are non-negative integers, is $ab-a-b$.

61. You have ten stacks of 50 pence pieces, nine stacks each containing 10 genuine coins and the other stack 10 counterfeit coins each weighing 1g more than a genuine one. If you know the weight of a genuine coin, identify the counterfeit stack with one weighing on a spring balance.

62. Show that

$$\int_0^\pi x\mathrm{f}(\sin x)\mathrm{d}x = \tfrac{1}{2}\pi \int_0^\pi \mathrm{f}(\sin x)\mathrm{d}x.$$

Let $\mathrm{f}(u) = u \sin^{-1} u \Rightarrow \mathrm{f}(\sin x) = x \sin x$

$$\int_0^\pi x \sin x \,\mathrm{d}x = \frac{\pi}{2}\int_0^\pi x \sin x \,\mathrm{d}x \Rightarrow \pi^2-4 = \tfrac{1}{2}\pi^2$$

Hence $\pi^2 = 8 \Rightarrow \pi = 2\sqrt{2}$.
Explain the fallacy.

63. The following celebrated problem first appeared in the *A.M.M.* in 1954.

We have to reconstruct the long division sum where all the digits except one have been replaced by x.

```
        xx8xx
   xxx)xxxxxxxx
       xxx
        xxxx
         xxx
         xxxx
         xxxx
         ....
```

64. We expect a biquadratic to have four roots over the complex field. Explain why we get only two roots on solving the biquadratic

$$x^2+xy-2y^2 = 1,\ x^2-3xy+2y^2 = 2.$$

65. In each of the following sequences u_n is defined by a recurrence relation and $u_1 = 1$:

(i) $u_{n+1} = u_n+\frac{1}{2}$, evaluate $\sum_{r=1}^{n} u_r$,

(ii) $u_{n+1} = u_n-u_n^2$, evaluate $\sum_{r=1}^{\infty} u_r^2$.

66. Prove $(n!)^2 > n^n$ for all integers $n > 2$.

67. A foreman noticed an inspector checking a circular hole of diameter 6 cm with two circular plugs of diameter 4 cm and 2 cm. He suggested that two more plugs be inserted to make certain that the fit was snug. If the new plugs are alike, find their diameter.

68. How many zeros are there at the end of the number 100!. Obtain an algorithm for the number of zeros at the end of n!.

69. Squares are described externally on the sides of a parallelogram. Prove that the centres of these squares are the vertices of another square. Would the proposition be simpler if the parallelogram was a rhombus?

70. Fermat's famous last theorem $x^n+y^n = z^n$ cannot be satisfied by non-zero integers if $n > 2$ can be proved quite easily in many special cases. Prove that three consecutive terms of a positive integer arithmetic progression can never be solutions of this equation. Can you find other special cases easy to prove—such as x, y, z cannot be the squares of odd integers?

71. If a, b, c, d are any four distinct integers prove that the product

$$(a-b)(a-c)(a-d)(b-c)(b-d)(c-d)$$

is divisible by 24.

72.* Choose a constant k so that the sum

$$\sum_{r=1}^{n} |k-a_r|$$

is to be a minimum, where $\{a_r\}$ is any sequence of monotonic increasing numbers $a_1 < a_2 < a_3 < \ldots < a_n$.

73.* Find constants b_0, b_1, b_2, b_3 so that a given cubic polynomial

$$a_0+a_1x+a_2x^2+a_3x^3 \equiv b_0+b_1\binom{x}{1}+b_2\binom{x}{2}+b_3\binom{x}{3},$$

where $\binom{x}{r} = \dfrac{x(x-1)(x-2)\ldots(x-r)}{r!}$.

Extend your method to show that the general polynomial

$$\sum_{r=0}^{n} a_r x^r$$

can be expressed *uniquely* in the form

$$\sum_{r=0}^{n} b_r\binom{x}{r}.$$

Hence deduce that the polynomial

$$\sum_{r=0}^{n} a_r x^r$$

takes on integer values for all integer values of x if and only if every coefficient b_r is an integer.

74. A moving staircase of n uniform steps visible at all times descends at constant speed. Two boys A and B move steadily down the staircase with A negotiating twice as many escalator steps per minute as B does. A reaches the bottom after taking 20 steps while B reaches the bottom after taking 15 steps. Find the value of n.

75. If 367^{654} were multiplied out, what will be the unit digit in the result?

76. Find the sum to infinity of the convergent series

$$\frac{1}{10}+\frac{2}{10^2}+\frac{3}{10^3}+\frac{4}{10^4}+\ldots$$

77. Three sailors share a pile of coconuts. The first sailor takes one more than half the pile; the second sailor takes one more than two-thirds of the remainder; the third sailor takes one more than three-quarters of the number still remaining. The nuts still left over were divided equally between them and left one over. What was the least number of nuts in the pile?

78. Any point P is chosen outside a sphere of radius r with centre O and a second sphere centre P and radius OP is

drawn. Show that the area of the second sphere lying inside the first is independent of the initial position of P. This result is not intuitive and should first be confirmed by considering the extreme cases.

79. If n is an integer >4 and *not* a prime number, prove that $(n-1)!$ is divisible by n. Does this help us to prove Wilson's theorem that, when n is an odd prime, $(n-1)!+1$ is divisible by n?

80.* Solve $\quad |x-1|+|x-2|+|x-3| = 3a,$

where a is a given positive number.

81. A square nut with external side of length a is to be unscrewed by a hexagonal spanner with internal side length b. Show that the ratio b/a must be between $\sqrt{(3/2)}$ and $(3-\sqrt{3})$.

82. Solve

(i) $x^2+x-12 = 0$

(ii) $|x^2|+|x| - 12 = 0$

(iii) $[x]^2+[x]-12 = 10$

where $|x|$ is the positive value of x and $[x]$ is the greatest integer $\leqslant x$.

83. Show that, when a circular disc of radius r rolls inside a circular rim of radius $2r$, the locus traced out by a point chosen arbitrarily on the edge of the disc is a diameter of the rim.

84.* If n boys of different height stand in a row, show there must be at least k boys standing in ascending order, or else in descending order of height, where k is the smallest integer $\geqslant\sqrt{n}$.

85. $ABCD$ is an $n\times m$ chessboard and a piece on it can move one square at a time horizontally or vertically. In how many ways can the piece move from one diagonal corner A to the opposite diagonal corner C, if the journey must be of minimum length?

86. Consider the four propositions:

(i) $\mathbf{p} \Rightarrow \mathbf{q}$.

(ii) $\mathbf{p} \Rightarrow \sim\mathbf{q}$.

(iii) $\sim\mathbf{p} \Rightarrow \mathbf{q}$.

(iv) $\sim\mathbf{p} \Rightarrow \sim\mathbf{q}$.

Which of these statements imply the truth of the propositions

(*a*) $\sim\mathbf{q} \Rightarrow \sim\mathbf{p}$.

(*b*) $\mathbf{p} \Leftrightarrow \mathbf{q}$.

(*c*) $\sim(\mathbf{p} \Rightarrow \mathbf{q})$?

87. A lattice point is defined to be a point whose coordinates are integers, zero admitted. Use the method of exhaustion to show that the number of lattice points on the boundary and inside the region bounded by the x-axis, the ordinates $x = 1$ and $x = 5$ and the curve $y = x^2$ is 60. Hence find a conjecture for the general case when $x = s$ is replaced by $x = n$, any positive integer, and prove your result by induction.

88. If $x+(1/x) = 1$, find the value of $x^{16}+(1/x^{16})$ and the values of $x^n+(1/x^n)$ for all integers n.

89. If $bc = a^p$, $ca = b^q$, $ab = c^r$, prove by two distinct methods that $pqr = p+q+r+2$.

90. Investigate inductively the following connected problems and prove your conjecture in each of the first three cases:

(i) Find the maximum number of segments into which a

straight line can be divided by n parts arbitrarily chosen on it.

(ii) Find the maximum number of regions into which a plane can be divided by n straight lines chosen arbitrarily.

(iii) Find the maximum number of sections into which a three-dimensional space can be divided by n arbitrarily chosen planes.

(iv) What will be the corresponding problem in four-dimensional space? Give a conjecture for the maximum number of sections.

91. How many squares can be seen on an ordinary chessboard? How many on an $n \times n$ chessboard? How many on an $m \times n$ rectangular chessboard? Give a proof in each case.

92. A knight is placed on the central square of a 5×5 chessboard. Find a journey in which the knight visits every square once and once only. Does your method (not, of course, trial and error) extend to boards of size 7×7 or 9×9? Can we generalize?

93. By considering the corresponding cases in one, two, and three dimensions, find conjectures for the following problems about four-dimensional regular polytopes. How many vertices, edges, faces, and solids are there in the four-dimensional regular tetrahedron, cube, and regular octahedron.

Verify that Euler's formula $n_0 - n_1 + n_2 - n_3 = 1$ is satisfied in each case and give a conjecture for the 4-dimensional content of each of these polytopes.

94.* Here are two problems where the inductive approach does not seem to suggest a method of solution, and yet the solution is evident once the right approach has been found:

(i) Find a polynomial $F(x)$ of degree n in x such that for

every integer $k \in [0,n]$ we have $F(k) = 2^k$. (Hint—think of permutations and combinations.)

(ii) The integers $(1,2,3,\ldots,n^2)$ are arranged in any order. Show that there will always be a subset of at *least n* numbers which are monotonic descending or else monotonic ascending. [Hint—relabel new numbers in mth position as (m_i, m_j).]

95. A well-known game with pennies is for two players A and B to place pennies in turn on a table such that no two pennies overlap nor does any penny extend over the edge of the table. The player who places the last penny is the winner. If A plays first, who wins and what is the winning strategy for the following tables:

(i) Rectangular table.

(ii) Circular table with concentric circle cut out.

(iii) Table in shape of equilateral triangle.

96. The same pennies game is adapted to three dimensions by using magnetized pennies and metal solids. Who wins and what is his winning strategy for the following solids:

(i) Regular tetrahedron.

(ii) Cube.

(iii) Prism with cross-section an equilateral triangle.

97. An interesting Cambridge game called 'Sprouts' starts with n spots marked on a sheet of paper. A move consists in drawing a line that joins one spot to another or to itself and then placing a new spot anywhere on the line. The rules are:

(i) The line may have any shape but it must not cross itself, nor cross a previously drawn line nor pass *through* a previously made spot.

(ii) No spot may have more than three lines emanating from it.

The winner is the last person to make a move. Decide which player will be the winner and show that a game must last for at least $2n$ moves and must finish in at most $(3n-1)$ moves.

98. Using the usual representation $[x]$ = the greatest integer $\leqslant x$, show that the graphs of $y = [x]$ and $y = [x+\frac{1}{2}]$ verify the result

$$[x]+[x+\tfrac{1}{2}] = [2x].$$

Test whether $\quad [x]+[x+\frac{1}{3}]+[x+\frac{2}{3}] = [3x].$

This suggests a general conjecture. State and prove this conjecture.

99. Given a pack of cards, a dice, and the properties of the number 142,857, make up a simple trick that would puzzle most audiences.

100.* An Icelandic patrol boat proceeding at 32 knots is overhauling a British trawler proceeding at 8 knots. When the boats are 5 nautical miles apart, a thick fog suddenly descends, whereupon the trawler changes course and proceeds in a new direction with her speed unchanged. Can the patrol boat be certain of intercepting the trawler although the new direction of the trawler is unknown to the patrol boat? If so, how should the patrol boat proceed?

Answers and Hints

1. Rational sides $2mn$, m^2-n^2, m^2+n^2 and all integers of form $5r$, $5r\pm1$, $5r\pm2$.

2. Common factor of A and B divides $A-xB$; $k = -3$.

3. Simplest to draw graphs of $y = \log_{10} x$ and $y = (\log_{10} 2)/x$.

4. Use Example 4 in Chapter 1; $\lambda = 12$, $\mu = -23$.

5. $N = \frac{1}{2}n(n-3)$; $\binom{N}{2}-\binom{n-3}{2}$ since $(n-3)$ diagonals from each vertex.

7. Six equations of type $(3+x)^3+(4-x)^3+(5+x)^3 = (6+x)^3$ (by suitable variation of the 4 signs) lead to *linear* equation for x; here $x = 3$, $\Rightarrow 6^3+1^3+8^3 = 9^3$.

8. (i) Difference of two squares.
   (ii) Quadratic by substitution $y = x^2+3x$.

9. Induction or binomial expansion.

10. (i) $\frac{1}{8}\pi^2$. (ii) $\frac{1}{12}\pi^2$.

11. $\frac{1}{3}n(n+1)(n+2)$; remember $\sum_1^r 1 = r$.

12. $x(1{,}000)+y(1) \equiv x+y \pmod{37}$.

13. Diagonal of parallelogram centre P whose sides lie along OA and OB. Use the inductive method and working backwards.

14. Use the Dirichlet-box principle.

15. Simplest proof by *geometry* from the usual triangle OAP with OA radius and AP tangent; join point where OP cuts circle to mid-point of AP.

16. Use trigonometry with $x = \tan\alpha$, etc. $\Rightarrow \alpha+\beta+\gamma = n\pi$ $\Rightarrow 2\alpha+2\beta+2\gamma = 2n\pi$ which gives required result.

17. $143 = \dfrac{1{,}001}{7} \Rightarrow abc\times 143 = \dfrac{abcabc}{7}.$

$1{,}667 = \dfrac{5{,}001}{3} \Rightarrow abc\times 1{,}667 = \left[\dfrac{abc0}{6}\right]\left[\dfrac{abc}{3}\right]$ with one-half of any remainder in first division carried on to the second division.

18. Straightforward!

19. This seems easiest done by trigonometry—let $AP = p$, $B\hat{A}P = \alpha$ and height $= h$, $\Rightarrow p_1+p_2+p_3 = p\sin\alpha+p\sin(A-\alpha)$ $+h-p\cos(\frac{1}{2}A-\alpha)$ and perpendicular from $Q = h-\frac{1}{2}p\cos(\frac{1}{2}A-\alpha)$. Equating these gives $\sin\alpha+\sin(A-\alpha)-\frac{1}{2}\cos(\frac{1}{2}A-\alpha)$ $= 0$ which is true for this triangle because $\sin A = \frac{1}{4}$.

20. $AB = 13$ cm; $\sin B = 12/13$ and $\sin C = 4/3$ with sine rule gives $AB+AC$.

21. Since a determinant has same value when transposed

$$\text{L.H.S.} = \begin{vmatrix} a & c & b \\ b & a & c \\ c & b & a \end{vmatrix}\begin{vmatrix} x & y & z \\ z & x & y \\ y & z & x \end{vmatrix} = \begin{vmatrix} A & C & B \\ B & A & C \\ C & B & A \end{vmatrix}$$

where $A = ax+by+cz$, $B = bx+cy+az$, $C = cx+ay+bz$.

22. Ordinary geometry properties do not seem to give a solution(!); coordinate geometry with origin O and chord AOB as x-axis will give solution, but working still involved. Should verify result by specialization.

23. Use scale 1:10 and let x and $(4-x)$ be the two parts of the ladder $\Rightarrow$ an equation

$$\frac{1}{x^2}+\frac{1}{(4-x)^2} = 1.$$

Now *insight* is needed to see that this equation can be put into an accessible form by putting

$$y = 2-x \Rightarrow \frac{1}{(2-y)^2}+\frac{1}{(2+y)^2} = 1.$$

Geometrically we are choosing length y = distance from centre of ladder to the point of contact with the box $\Rightarrow y = (5-\sqrt{17})^{\frac{1}{2}}$ and height $= 37{\cdot}6$ ft.

24. Straightforward manipulation.

25. See Example 4 in Chapter 2; when

$$\text{R.H.S.} = 0,\ I_n = (An+B)\,.\,2^n$$

for arbitrary constants A and B; when R.H.S. $= n$, we try $I_n = an+b$; see the anology with linear differential equations of second order.

26. (i) Use sine rule $a = k\sin A$, etc.

(ii) Use geometry by producing BA to D where $AD = AC$ and considering circumcircle of $\Delta ACD \Rightarrow$ easy to show BC is a tangent, whence result follows.

27. Use inductive process for $1D$, $2D$, $3D$ to see that we can write them in the form $n_0-n_1+n_2-n_3 = 1$ (where n_0 = number of edges, etc.). The extension to $4D$ and nD is now obvious.

28. Specialization convinces us of the truth of the result, but it is quite awkward to put the proof in a form to cover arbitrary numbers a, b, c.

29. Use inductive procedure to conjecture number is $(n+1)^2$; prove by considering $y+z = 2k$ where $0 \leqslant k \leqslant n$.

30. Remember that two sides of a triangle are together greater than the third.

31. (i) No lines intersect inside circle $\Rightarrow (n+1)$ regions.

(ii) All lines concurrent at same point $\Rightarrow 2n$ regions.

(iii) Use inductive process to obtain conjecture that

$$N = 1+\frac{n(n+1)}{2}$$

and prove by induction.

32. Number is 3^{27} and *not* $(27)^3$; take logs to evaluate the index first and repeat to evaluate characteristic of the log of the number.

33. Common sense; $2^{10} = 1{,}024 \sim 10^3$ and $3^9 = 19{,}683 \sim 20{,}000 = 2\times 10^4 \Rightarrow 2^{85} > 32\times 10^{24}$ and $3^{53} < \frac{64}{3}\times 10^{24} \Rightarrow 2^{85}$ is arger.

34. $n \equiv 0 \pmod 7$ and $\equiv m-1$ (mod $m = 2, 3, 4, 5, 6$) $\Rightarrow n \equiv -1$ (mod 2, 3, 4, 5, 6) $\Rightarrow n = 60k-1$ (since 60 is L.C.M. of 2, 3, 4, 5, 6) $\Rightarrow n = 119$.

35. $l = -1$ and $\sum_{r=1}^{n} u_r = \log\{n!/(n+1)^n\}$.

36. *Necessity:* let $a_1, a_2 \in A$ and $f(a_1) = f(a_2) \Rightarrow gf(a_1) = gf(a_2) \Rightarrow I_A(a_1) = I_A(a_2) \Rightarrow a_1 = a_2$. Hence f is one–one.

Let $b \in B$ and $g(b) = a \in A \Rightarrow f(a) = fg(b) = I_B(b) = b$. Hence f is surjective or onto. Hence f is bijective.

Sufficiency: given $f: A \to B$ is one–one and onto, define $g: B \to A$ by $g(b) = a$ when $f(a) = d$. Similar proof to necessity, but in reverse.

37. Of the various ways of unfolding net of room, one way gives distance 40 ft and all the others are longer.

38. Unwrap surface of hill to sector of circle radius 1,950 ft and arc $1{,}500\pi$ ft; least distance is $1{,}900 \sin(5\pi/13)$ ft.

39. $(x^4+qx^2+p) = (x^2+px+a)(x^2-px+p/a) \Rightarrow p = 0, q = a$ or $p = a^2$, $q = 2a-a^4$.

40. (i) $n!$

(ii) $n(n+1)/2$.

(iii) $\left[\frac{n+1}{2}\right]$.

41. (i) None, since $|\text{L.H.S.}| \leqslant 5$.

(ii) $(2n+1)\pi$.

(iii) $2\pi\theta = 2n\pi \pm \left(\frac{\pi}{2}-\pi\theta\right) \Rightarrow \theta = 2n-\frac{1}{2}$ or $\frac{2n+1}{6}$.

(iv) all values of θ, since an identity.

42. Point (1, 1); area of $\Delta = 2k^2/(k^2-1)$, minimum $= 2/\left(1-\frac{1}{k^2}\right) \leqslant 2$.

43. Consider suitable sections of the square pyramid whose vertices are the five centres.

44. Simple arithmetic (modulo 7) allowing for the leap years.

45. Diophantine problem with unique solution 80 hens, 19 pigs, 1 cow.

46. Either use inductive process to conjecture $f(n) = \frac{1}{2}n(3-n)$ or write $x = 1$, $y = r$ for $r = 1$ to $(n-1)$ and add.

47. Curved area of cap of sphere thickness $h = 2\pi Rh$; $74\frac{1}{2}\pi$; spheres would cut orthogonally.

48. The total number of handshakes is even.

49. The other point $(-1, 0)$ is given by $t = \infty$.

50. $x = 1$, $y = 4$ and $x = 1{,}716 \equiv 5 \pmod{29}$, $y = 2$; *no*.

51. Specialization suggests the arc PQ is the arc of a circle cutting the arms at right angles; a proof would require a knowledge of isoperimetric theorems and methods.

52. 14 days.

53. (i) 5×10^9.

(ii) b or $1/b$.

54. $a = 2(\Delta \tan \frac{1}{2}\alpha)^{1/2}$, $b = c = (2\Delta \operatorname{cosec} \alpha)^{1/2}$—show Δ with given base and given angle opposite has greatest area when it is isosceles.

55. Use Dirichlet's box principle.

56. (i) Specializing gives 25, 41, and 5,041.

(ii) $6! \times 7! = 10!$

57. Six; each square is reduced in ratio $1:\sqrt{2}$.

58. $5\frac{1}{16}$ gallons.

59. Specialization gives conjecture 6; ignore all but last two digits of a^2.

60. Use Euclid's algorithm.

61. Take r coins from rth stack for $r = 1, \ldots, 10$; excess weight of n g identifies nth stack as the false one.

62. f(sin x) becomes algebraic when we substitute $x = \sin^{-1} u$.

63. Quotient is clearly $x080x$, first digit in dividend is 1, etc.

64. Consider their representation in coordinate geometry—we have two hyperbolae with one asymptote common to both. Hence other roots lie at infinity in direction $y = x$.

65. (i) Simple to show $u_n = \frac{1}{2}(n+1) \Rightarrow \sum_1^n u_r = \frac{1}{4}n(n+3)$,

(ii) we cannot find u_n here, but we use $u_n^2 = u_n - u_{n+1}$

$\Rightarrow \sum_1^n u_r^2 = u - u_{n+1} \to u_1$ as $n \to \infty$, by showing that $u_n \to 0$.

66. Write $(n!)^2 = (1 \,.\, n)(2 \,.\, n-1)(3 \,.\, n-2) \ldots (n \,.\, 1)$.

67. $2\frac{3}{7}$ cm; use perimeter and areas of the triangles of centres.

68. Inductive method gives 24 and a clarification of the method suggests the algorithm;

$$\text{number of zeros} = \sum_{r=1}^{k} n_r$$

where

$$n_1 = \left[\frac{n}{5}\right] \text{ and } n_{r+1} = \left[\frac{n_r}{5}\right] \text{ and } n_{k+1} = 0.$$

69. For $ABCD$ parallelogram ordinary geometry seems best; but for rhombus we have a simple proof because the figure is invariant under the group of reflections about a diagonal of rhombus and under the group of rotations through $\frac{1}{2}\pi$ about its centre O.

70. $(a-d)^n + a^n = (a+d)^n \Rightarrow (k-1)^n + k^n = (k+1)^n$ must have rational roots etc. Use inductive process first for case $n = 3$. For second choice, odd squares are of form $8N+1 \Rightarrow x, y, z$ cannot be three odd squares.

71. Dirichlet box-principle; residues mod 3 = 0, 1, 2 $\Rightarrow$ at least two of a, b, c, d have same residue (mod 3) $\Rightarrow$ product divisible by 3, etc.

72. Inductive investigation suggests conjecture $k = a_{r+1}$ when $n = 2r+1$ and k any value $a_r < k < a_{r+1}$ when $n = 2r$.

73. $b_0 = a_0$; $b_1 = a_1 + a_2 + a_3$; $b_2 = 2!a_2 + 3!a_3$; $b_3 = 3!a_3$; prove that $\binom{x}{r}$ is an integer when x is an integer.

74. 30 steps.

75. Ends in 9 since 7^4 ends in 1.

76. $S_n = \sum_1^\infty \frac{1}{10^r} + \sum_2^\infty \frac{1}{10^r} + \sum_3^\infty \frac{1}{10^r} + \ldots = \frac{1}{9} + \frac{1}{90} + \frac{1}{900} + \ldots = \frac{10}{81}$.

77. 128 nuts.

78. Area $= \pi r^2$.

79. Let $n = p \times q$ where p is the largest prime divisor; consider separately the cases $p \neq q$ and $p = q$; *no.*

80. Consider separately the intervals $(-\infty, 1]$, $(1, 2]$, $(2, 3]$, and $(3, \infty]$. No solution $a < \frac{2}{3}$, $x = 3a$ and $4-3a$ if $\frac{2}{3} < a < 1$, $x = 2 \pm a$ if $a \geqslant 1$.

81. The diagonal of the square must be greater than the distance between the opposite sides of the hexagon.

82. (i) -4, 3.
(ii) ± 3.
(iii) $-3 \leqslant x < -2$ and $3 \leqslant x < 4$.

83. Simple geometry.

84. Relabel each boy (x, y) where x is the largest subset in ascending order (including himself) to his left and y is the largest subset in descending order (including himself) to the right; show every boy has a unique $(x, y) \Rightarrow n \leqslant k^2 \Rightarrow k$ is smallest integer greater than or equal to $\sqrt{n}$.

85. Isomorphic with problem of arranging $(n-1)$ pence and $(m-1)$ shillings in a row $\Rightarrow (n+m-2)!/(n-1)!(m-1)!$

86. (*a*) prop (i), (*b*) prop (i)+prop (iv), (*c*) any except prop (i).

87. Inductive investigation conjectures

$$1+\frac{n}{2}(n+1) = (n^2+n+2)/2.$$

88. Let $S_n = x^n + \dfrac{1}{x^n}$; $S_{2n} = (S_n)^2 - 2 \Rightarrow S_{16} = -1$; inductive investigation suggests conjecture $S_{n+3} = -S_n$; prove this $\Rightarrow S_{3r+1} = (-1)^r S_1 = (-1)^r$, $S_{3r+2} = (-1)^{r+1}$, $S_{3r} = (-2)^r$.

89. (i) Use indices from $a^{p+1} = b^{q+1} = c^{r+1}$.
(ii) Take logarithms.

90. (i) $1+n \equiv 1+\binom{n}{1}$.

(ii) $\frac{1}{2}(n^2+n+2) \equiv 1+\binom{n}{1}+\binom{n}{2}$.

(iii) $1+\binom{n}{1}+\binom{n}{2}+\binom{n}{3}$ all probable by induction or direct from observation $N^r_{n+1} - N^r_n = N^{r-1}_n$; conjecture $1+\binom{n}{1}+\binom{n}{2}+\binom{n}{3}+\binom{n}{4}$.

91. $\sum\limits_1^8 r^2 = 204$; $S_n = \frac{n}{6}(n+1)(2n+1)$; show $S_{n+r+1} = S_{n+} + \frac{1}{2}n(n+1)$ and hence deduce $S_{m,n} = \frac{1}{6}(n+1)(3m-n+1)$.

92. Start at any diagonal corner and keep moving clockwise around the centre moving each time as far away from the centre as possible; the method extends to square boards of side $4n+1$ but not of side $4n-1$.

93. Pattern apparent from a table for each solid; tetrahedron $V = 5$, $E = 10$, $F = 10$, $S = 5$; cube $V = 16$, $E = 32$, $F = 24$, $S = 8$, octahedron $V = 8$, $E = 24$, $F = 32$, $S = 16$.

94. (i) Think of permutations and combinations—total number of ways of choosing any number of things from n different things is 2^n (take it or leave it) and $= \binom{n}{0}+\binom{n}{1}+\ldots+\binom{n}{n}$. Now easy to see that

$$F(x) = \binom{x}{0}+\binom{x}{1}+\binom{x}{1}+\ldots+\binom{x}{n}.$$

(ii) Think of increasing and decreasing as two components of the same number and relabel each number, say m, as (m_i, m_d) according to the largest number m_i increasing in front of it (including m) and m_d as decreasing behind it (including m).

95. (i) A wins by starting in centre and always playing the reflection in the centre of any move B makes.

(ii) B wins by playing the reflection in the centre of A's move.

(iii) A conjecture is that A wins by starting in the centre of the triangle.

96. (i) Number faces 1, 2, 3, 4; B wins by copying A's move on face r on the face $(r+1)$, or $(r-1)$ if $(r+1)$ place already filled.

(ii) B wins by copying any move A makes with same move on the opposite face.

(iii) A wins by starting at centre of a sloping face and then ① if B moves on an end face, A copies move on opposite face ② if B moves on one of a different sloping face, he copies move on the other sloping face ③ if B moves on same face as A started on, A makes the usual symmetrical move with respect to the centre of that face.

97. A wins if n odd (>1), B wins if n even or $n=1$.

98. $[x]+\left[x+\frac{1}{n}\right]+\left[x+\frac{2}{n}\right]+\ldots+\left[x+\frac{n-1}{n}\right]=[nx]$ for all integers n; prove by induction.

99. Pre-arrange the pack so that the top six cards are in the sequence Ace, 4, 2, 8, 5, 7 in various suits. False shuffle, and then take off these six cards one by one, asking a spectator to write down the corresponding digits (142857) on the board (or a slate) as you do so. Ask him to throw the dice, and quickly cut your set of six cards at the appropriate place (to be learned beforehand). Give this set face down to the spectator. Then multiply the number on the board by the number thrown on the dice. When finished, ask the spectator to call out the cards in his hand, one by one in sequence from the top. The order should agree with the quotient on the board.

100. Surprisingly, the answer is *yes*. The patrol boat carries on for 4 nautical miles and then sets off with speed unchanged on a logarithmic spiral whose origin is the point where the trawler vanished and such that *the patrol's boat distance from the origin increases at the same rate* as the trawler's (i.e. at 8 knots). The two courses must therefore intersect before the patrol boat has made one complete circuit of the origin.

Suggestions for Further Reading

1. BOOKS ON PROBLEM-SOLVING

G. Polya, *How to Solve It* (Doubleday, 1957)
Mathematics and Plausible Reasoning (O.U.P., 1954)
Mathematical Discovery, 2 vols. (Wiley, 1962)

2. SOME BOOKS ILLUSTRATING THE SOLVING OF PROBLEMS

D. Pedoe, *The Gentle Art of Mathematics* (Penguin Books, 1959)
S. G. Ogilvy, *Through the Mathescope* (O.U.P., 1956)
M. Kac and S. M. Ulam, *Mathematics and Logic* (Penguin Books, 1971)
G. Gamor and M. Stern, *Puzzle Math* (Macmillan, 1960)
H. Steinhaus, *Mathematical Snapshots* (O.U.P., 1969)

3. SOME BOOKS ON PROBLEMS

Shlarsky, Chentsor and Yaglom, *The Olympiad Problem Book* (Freeman, 1962) (Contains problems of elementary nature but unusual type; some are difficult)
The Stanford University Competitive Examinations in Mathematics
C. T. Salkind (ed.), *The M.A.A. Problem Books* (Random House, 1966)
E. Rapaport (tr.), *The Hungarian Problem Book* (Random House, 1963)

Transworld Student Library

General Editor:

H. GRAHAM FLEGG, M.A., D.C.Ae., C.Eng., F.I.M.A., M.I.E.E., A.F.R.Ae.S., F.R.Met.S.
Reader in Mathematics, The Open University

Transworld Student Library is a series of paperback books devoted to topics in Mathematics and the Sciences designed to meet the modern educational needs of the independent learner. The academic level varies with the familiarity of the material covered, from general introductory presentations suitable for sixth-form pupils and school leavers to more specialized topics treated at first- or second-year University level. At the same time, many of the books will be of interest to the general reader with an enquiring mind who wishes to become acquainted with some of the more recent developments in Mathematics and the Sciences.

For the most part, the treatment breaks away from the traditional presentation geared to the classroom situation and provides a refreshingly new approach to the subject matter being discussed.

The criteria for inclusion in the series are that a book shall be clear and understandable in what it has to say; that the academic standard shall be impeccable; that the author shall be genuinely enthusiastic in his desire and ability to communicate to the reader; that the presentation shall be as concise as is compatible with clear understanding; that particular attention shall be paid to the provision of illustrative material; and that the principal aim shall be to stimulate the reader's interest and his desire to study further as well as to provide information and general background material.

Some of the books in the Library are first English translations of outstanding foreign works where these meet the criteria for inclusion, but the majority are specially commissioned from authors who, whilst being specialists in their subjects, are nevertheless prepared to break away from traditional and now outmoded approaches and present their material in a manner consistent with the new adult educational requirements.

THEORETICAL STATISTICS—BASIC IDEAS

by STANLEY N. COLLINGS
Reader in Statistics, The Open University

Theoretical Statistics does not pretend that statistics is a non-mathematical subject. Starting from a few A-level concepts, and confining itself to discrete situations, it provides a clear introduction to how these concepts are used in formulating the basic ideas upon which probability and sampling notions are built.

0 552 40002 5 100 pages **70p** T121

BOOLEAN ALGEBRA

by H. GRAHAM FLEGG
Reader in Mathematics, The Open University

Boolean Algebra provides a general introduction from first principles to the algebra of two—state devices through a discussion of sets, propositions and simple switching circuits. The algebra presented here now forms part of most 'modern' mathematics syllabuses and is of considerable importance in a number of fields of application.

0 552 40001 7 160 pages **80p** T122

POINTS AND ARROWS: THE THEORY OF GRAPHS

by ARNOLD KAUFMANN
Professor at L'Institut polytechnique de Grenoble

Translated by H. Graham Flegg

Points and Arrows provides a sound elementary introduction to the theory of graphs, a branch of modern mathematics of increasing importance in various sciences, sociology, economics and business studies. Applications to various optimal path problems are discussed, and a fascinating glimpse is provided into recent theories of pattern recognition systems.

0 552 40003 3 160 pages **80p** T123

BASIC MATHEMATICAL STRUCTURES I

by NORMAN GOWAR and GRAHAM FLEGG

This book introduces some of the ideas which underly almost all abstract algebra. The basic ideas of mathematical structure are introduced in chapters on 'sets', 'binary operations', and 'relations'. Examples of various structures are then given, starting from very simple ones and building to more complicated cases. This is not intended to be a detailed exposition of any particular algebraic structure, but various structures are compared and, it is hoped, some feeling of their relationships one to another developed.

The book should be suitable for sixth-form pupils and teachers of sixth-form pupils, as well as for first-year undergraduates. It should also be of interest to anybody who wishes to find out some of the 'flavour' of algebra but who does not want to tackle a detailed and specialised text.

552 40011 4 **80p** T166

TOWARDS QUANTUM MECHANICS

by DR. J. CUNNINGHAM

Through a study of a single physical problem—the harmonic oscillator—this book provides an elementary introduction to the underlying concepts of quantum theory. Starting from the classical theory of oscillations, the need for a quantum theory to replace the classical theory of the very small is traced historically, and the mathematical basis of the quantal model is developed in detail. Mathematics to 'A' level, with some knowledge of elementary mechanics and calculus, is all that is asked of the reader.

552 40010 6 **75p** T167

NUMERICAL ANALYSIS

by DR. A. GRAHAM

The techniques of Numerical Analysis have become of great importance in the development of mathematical modelling with the aid of computers. This book discusses a wide range of techniques, suggests their advantages and disadvantages, and accompanies the majority of them with some form o error analysis. Throughout, the mathematics has been kep as simple as is compatible with demonstrating a number o important results and properties of the various techniques.

552 40013 0 **75p** T16

EVOLUTION OF MATHEMATICAL CONCEPTS

by RAYMOND L. WILDER

This book looks at mathematics as a cultural component of society, and studies its evolution from an anthropological stand-point. The basic concepts of number and geometry are used to illustrate how and why mathematics as a whole has developed the way that it has. A fascinating book for mathematicians and non-mathematicians alike, providing a real insight into *what mathematics really is* without demanding any familiarity with the technicalities of the subject.

552 40018 5 **85p** T169

INTRODUCING REAL ANALYSIS

by DAVID FOWLER

Lecturer in Mathematics and Manager of the Mathematics Research Centre, University of Warwick

A genuinely introductory book that makes few assumptions of the reader's knowledge other than an intuitive acquaintance with functions and rational numbers, and some experience in manipulating derivatives of elementary functions. Great care is taken that the reader is not overwhelmed by the notation required by the subject, the notation being evolved as the book progresses. Written with clarity and humour, it provides an exemplary introduction to a subject that forms part of the stock-in-trade of every mathematician.

0 522 40017 3 128 pages **75p** T78

NEW PERSPECTIVES IN EVOLUTION

by HANS QUERNER, HELMUT HOLDER, ALBRECHT EGELHAAF, JURGEN JACOBS, GERHARD HEBERER

New Perspectives in Evolution is an authoritative and comprehensive survey of the present state of knowledge of the origin and evolution of living things, and of man in particular. It places Darwin's theory in its historical context, and shows how a variety of sciences, themselves greatly developed since Darwin's day, have confirmed and refined the main elements of the theory of evolution.

This book has been written by leading experts in these various fields. Addressed to teachers of biology, biology sixth-formers and undergraduates, as well as the interested laymen it is readable, authoritative and profusely illustrated, and offers a fascinatingly diverse picture of the origin and evolution of living things.

0 552 40012 2 160 pages **80p** T79